Evaluating Impacts of Peacebuilding Interventions

Approaches and methods, challenges and considerations

Principal Authors
Diana Chigas, CDA
Madeline Church, Saferworld
Vanessa Corlazzoli, Search for Common Ground

Date:
May 2014

Practice Products for the CCVRI

Improving Measurement in DFID Crime, Conflict & Violence Programming

This document is one of a series of Practice Products developed under the *Conflict, Crime, and Violence Results Initiative (CCVRI)*. The full set of products is intended to support DFID country offices and their partners to develop better measures of programme results in difficult conflict and fragile environments.

DFID recognises the need to focus on the *results* of its work in developing countries. To this end, DFID strives to account better for our efforts on behalf of UK taxpayers, offering clarity regarding the value and impact of our work. The Results Initiative operates under the assumption that we will achieve our development objectives with our national partners more effectively if we generate—collectively—a clear picture of the progress being made.

Within DFID, the Conflict Humanitarian and Security Department has established a partnership with a consortium of leading organisations in the fields of conflict, security and justice to develop more effective approaches to the use of data in the design, implementation and evaluation of programmes that contribute to reducing conflict, crime and violence.

In addition to producing these Practice Products, the consortium has established a Help Desk function to provide direct and customised support to country offices as they endeavour to improve measurement of results in local contexts.

The Help Desk can be accessed by contacting helpdesk@smallarmssurvey.org.

The views expressed in this Practice Product are the sole opinions of the authors and do not necessarily reflect the opinions of all consortia partners. This Practice Product does not reflect an official DFID position.

Members of the consortium

Document Summary

Title:

Evaluating Impacts of Peacebuilding Interventions: Approaches and methods, challenges and considerations

Purpose and intended use of this document:

DfID's 2009 Peacebuilding and Statebuilding Framework outlined an integrated approach with statebuilding and peacebuilding at the centre of its work in fragile and conflict-affected countries (FCAS). Yet there is relatively little evidence of the peacebuilding impacts of DfID's engagement in these situations, and there are serious challenges to evaluating them. This guidance is intended to help staff a) understand the challenges and opportunities of evaluating impacts of interventions that aim to build and achieve peace by targeting, directly or indirectly, key drivers of conflict and peace (peacebuilding interventions), and b) work with evaluators to ensure evaluation designs are appropriate, robust and conflict-sensitive. This guidance builds on a study already done for DFID on *Broadening the Range of Designs and Methods for Impact Evaluations* (which addresses impact evaluation (IE) of development interventions generally) and introduces a peacebuilding lens.

Key questions this document addresses:

- What does impact evaluation mean in a peacebuilding context?
- What are the challenges to evaluating impacts of peacebuilding interventions?
- How is evaluating impacts of peacebuilding interventions different from evaluating impacts of development interventions?
- What are key considerations in the design of impact evaluations of peacebuilding interventions?
- What kinds of evaluation questions are useful for peacebuilding interventions?
- What evaluation approaches are useful for evaluating impacts of peacebuilding interventions?
- What are the strengths and weaknesses of different evaluation approaches and methodologies for impact evaluation of peacebuilding—experimental and quasi experimental approaches, theory-based and case study-based designs and participatory approaches to impact evaluation?
- How does the evaluation process interact with the conflict context, and how can conflict sensitivity of the evaluation process be ensured?

Key messages/essential 'take aways':

- Impact evaluation is interpreted in different ways in peacebuilding: a) attribution of specific outcomes to an intervention; b) assessment of effects of the intervention on broader drivers of conflict; c) assessment of whether and how effects have been sustained over the longer term. It is important to define and agree on what is meant by impact among country offices, implementers, key stakeholders (and ultimately evaluators) to avoid misunderstandings.
- For peacebuilding, assessing an intervention's contribution to addressing drivers of conflict is important and distinguishes impact evaluation of peacebuilding from impact evaluation of development or other interventions
- It is important to decide explicitly what *level* of impact the evaluation will focus on: project/programme level objectives, broader effects on drivers of conflict beyond the participants, or longer term and sustained impacts either at project/participant level or beyond. This will be important for choosing an appropriate evaluation design.
- *Conflict analysis* is critical to evaluation of peacebuilding impact to understand and test the *relevance* of the intervention to the conflict drivers. For evaluating peacebuilding impacts, it is not enough to demonstrate that the intervention has contributed to changes in the *context*; the intervention must be relevant to those

aspects of the context of conflict or fragility that directly shape or influence how the conflict develops and whether violence occurs.

- Articulating *theories of change* of peacebuilding interventions is critical to evaluating impacts of peacebuilding. Whatever evaluation approach or design is chosen, it is important to understand the assumptions (theories) about how the intervention is expected to produce or contribute to desired changes in the conflict, so that they can be tested and verified (or not) in the evaluation.
- Because fragile and conflict-affected contexts are complex, it is important to examine an intervention from the 'inside out' (i.e. looking at the intervention's activities, outputs, outcomes and theory(ies) of change) as well as the 'outside in' (understanding the evolution of the conflict dynamics in the issues or area the intervention is trying to influence).
- In evaluating impacts of peacebuilding, attribution of changes to an intervention is not possible, because changes in drivers of conflict and fragility come about because of the combination of many factors—including other interventions, contextual (including regional and global) developments—often working together. Examining contribution is more realistic and productive than trying to attribute changes to an intervention.
- There is no one best evaluation approach for peacebuilding—the approach should be matched to the purpose of the evaluation and the evaluation questions of interest. Nevertheless, experimental approaches (e.g. randomised control trials, or RCTs) and quasi-experimental approaches generally will have very limited utility because they cannot answer many important impact evaluation questions regarding peacebuilding and because methodological challenges and conflict sensitivity concerns make them difficult and risky to implement in conflict contexts.
- Integration of participatory approaches into evaluation designs is especially useful in evaluations of peacebuilding impacts because they allow for more nuanced understanding of the changes in the conflict contexts and for the incorporation of differing perspectives on conflict and peace (inherent in conflict situations) into the evaluation process.
- No evaluation approach and no single method are without gaps or weaknesses. Robust evaluations of peacebuilding impacts mix methods (qualitative and quantitative) and mix, where possible and appropriate, approaches (theory-based, case-based, participatory, experimental/quasi-experimental) in order to fill gaps and compensate for weaknesses.
- There are many ways an evaluation process interacts negatively with the conflict context. Conflict can affect access to data, security for evaluators and participants and undermine the conditions necessary for the evaluation process to yield valid findings. The evaluation process can also affect the conflict dynamics in a number of ways. The mere presence of foreigners may increase security risks for local people. Choice and implementation of methods—for example how questions are asked, who is included and who is not included in the evaluation process—can heighten trauma, increase frustration and salience with unmet expectations, and exclude important voices. Evaluation processes can send negative implicit ethical messages of disrespect, disempowerment and lack of agency.

Intended audience of this document (including assumed skill level):

DFID advisers, implementing partners who are considering, planning, commissioning and managing impact evaluations of peacebuilding interventions or integrated peacebuilding and development interventions in fragile and conflict affected states. A basic knowledge of peacebuilding and development programming design, monitoring and evaluation is assumed.

Key topics/tags:

Impact evaluation
Peacebuilding
Monitoring and evaluation
Conflict sensitivity
Evaluation approaches
Experimental and quasi-experimental evaluation designs
Randomised control trials (RCTs)
Theory-based evaluation

Case studies
Participatory impact assessment
Mixed methods
Conflict analysis

Authors and their organisations:

Diana Chigas, CDA Collaborative Learning Projects
Madeline Church, Saferworld
Vanessa Corlazzoli, Search for Common Ground

Reviewers: Rachel Goldwyn, independent consultant; Carlisle Levine, BLE Solutions, LLC.

Cross-references to other documents in the series:

Bayne, S. with Vaux, T. (2013) *Integrated development and peacebuilding programming: Design, monitoring and evaluation* (CARE and CDA)

Corlazzoli, V., and White, J. (2013) *Practical Approaches to Theories of Change in Conflict, Security and Justice Programmes: Part II: Using Theories of Change in Monitoring and Evaluation* (Search for Common Ground)

Corlazzoli, V., and White, J. (2013) *Back to Basics: A Compilation of Best Practices in Design, Monitoring and Evaluation in Conflict and Fragile Environments* (Search for Common Ground)

Corlazzoli, V., and White, J. (2013) *Measuring the Unmeasurable: Solutions to Measurement Challenges in Conflict and Fragile Environments* (Search for Common Ground)

Goldwyn, R. and Chigas, D. (2013) *Monitoring and evaluating conflict sensitivity: Methodological challenges and practical solutions* (CARE and CDA)

Small Arms Survey (2013) *Tools for measurement, monitoring and evaluation: Making conflict, crime and violence data usable* (SAS)

Small Arms Survey (2013) *Tools for measurement, monitoring and evaluation: In-depth focus on surveys*

Woodrow, P. and Oatley, N. (2013) *Practical Approaches to Theories of Change in Conflict, Security & Justice Programmes: Part I: What they are, different types and how to develop and use them* (CDA Collaborative Learning Projects)

Table of Contents

List of Tables

List of Boxes

1   Introduction

DfID's 2009 Peacebuilding and Statebuilding Framework outlined an integrated approach with statebuilding and peacebuilding at the centre of its work in fragile and conflict-affected countries (FCAS)[1]. The framework centres on addressing the causes and effects of conflict and fragility, and building conflict resolution mechanisms, along with supporting inclusive political settlements and processes, developing core state functions and responding to public expectations. In 2012 DfID reaffirmed this strategy, noting the 'wide recognition that poverty reduction can only be achieved if we address conflict and fragility through a peacebuilding and statebuilding approach' and calling for *all* interventions in FCAS to contribute to tackling conflict and fragility[2].

In this context, it is critical to understand whether, how and why interventions in FCAS have contributed to addressing drivers of conflict and fragility. Yet as the same How To Note on *Results in Fragile and Conflict-Affected States and Situations* underlines, there is relatively little evidence of the peacebuilding impacts of DfID's engagement in these situations, and there are serious challenges to evaluating them[3]. Evidence of what has worked, what has not worked, and why is needed to improve country strategies, intervention design and implementation and to fulfil responsibilities of accountability to participants in programmes, host governments and UK taxpayers.

This guidance is intended to help staff a) understand the challenges and opportunities of evaluating impacts of interventions that aim to build and achieve peace by targeting, directly or indirectly, key drivers of conflict and peace (peacebuilding interventions)[4] and b) work with evaluators to ensure evaluation designs are appropriate, robust and conflict-sensitive. This guidance builds on a study already done for DfID on *Broadening the Range of Designs and Methods for Impact Evaluations* (which addresses impact evaluation (IE) of development interventions generally[5]) and introduces a peacebuilding lens.

Section 2 discusses what impact evaluation means in a peacebuilding context. It highlights misunderstandings that have resulted from different conceptions of 'impact evaluation' by professional evaluators, on the one hand, and programmatic and policy staff and implementers on the other, in

---

[1] DfID (2010) *Building Peaceful States and Societies: A DFID Practice Paper*. London: DFID. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67694/Building-peaceful-states-and-societies.pdf.

[2] DfID (2012) *How To Note: Results in Fragile and Conflict-Affected States and Situations*. London: DFID. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67437/managing-results-conflict-affected-fragile-states.pdf.

[3] *Id*., para. 9.

[4] This paper distinguishes peacebuilding from conflict sensitivity. Conflict sensitive development seeks to ensure that design, implementation and outcomes of development interventions do not undermine peace or exacerbate conflict, and contribute to peace where possible. Peacebuilding aims to build peace by targeting and addressing key drivers of conflict. *See* Bayne, S. with Vaux, A. (2013) *Integrated development and peacebuilding programming: Design, monitoring and evaluation*. CCVRI Guidance Product. London: DfID; Goldwyn, R. & Chigas, D. (2013) *Monitoring and evaluating conflict sensitivity: Methodological challenges and practical solutions*. CCVRI Guidance Product. London: DfID.

[5] Stern, E. Stame, N., Mayne, J., Forss, K., Davies, R. & Befani, B. (2012) *Broadening the Range of Designs and Methods for Impact Evaluations: Report of a study commissioned by the Department for International Development*. London: DfID. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf.

order to develop a common understanding of what impact evaluation entails. *Section 3* provides general considerations about impact evaluation designs of peacebuilding interventions or peacebuilding effects of development interventions. *Sections 4-7* outline different methodologies that are commonly used in impact evaluations to address the 'attribution problem' of connecting the intervention to the results of interest, and discusses the strengths, weaknesses and challenges of each in peacebuilding contexts, as a basis for exploring the appropriateness and feasibility of designs and methods. A comparative summary of the strengths and limitations of the approaches is provided in **Annex 1** to provide a user-friendly tool for choosing appropriate methodology given the nature of the intervention, the level of impacts of interest and the evaluation purpose and questions posed. Resources for further information on the different approaches are provided in **Annex 2**.

2 Understanding Impact Evaluation in Peacebuilding Contexts

2.1 What is 'impact' and what is 'impact evaluation'?

As a term, 'impact' has been used in different ways in programme design and evaluation both within DfID and among policy makers, programming staff and professional evaluators. Within DfID alone there are several ways in which the term 'impact' is used:

- For programme and project planning, DfID's 2011 Guidance on using the new logical framework refers to impact as something 'not intended to be achieved solely by the project…a higher-level situation that the project will contribute towards achieving'[6]. For development projects, the Impact normally focuses on achieving one of the Millennium Development Goals, or Climate Change targets. For humanitarian projects, the Impact might be the restoration of peace and security, or the maintenance of basic services[7]. The OECD DAC's 2012 guidance on *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility* also focuses on higher-level objectives, or 'results that lie beyond the immediate context of an intervention and influence the intensity, shape or likelihood of a conflict'[8].

- For evaluation, DfID's evaluation policy adopts the OECD DAC's (2002) definition of impact as 'positive or negative, primary and secondary effects produced by an intervention, directly or indirectly, intended or unintended'[9].

- The How To Note on *Results in Fragile and Conflict-Affected States and Situations* introduces a methodological lens to conception of impact, defining it as 'the outcome of an intervention compared to what it would have been in the absence of the intervention'[10].

---

[6] DfID (2011) *How to note: Guidance on using the revised Logical Framework*. London: DfID. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/253889/using-revised-logical-framework-external.pdf. p. 13.

[7] *Id.*, p. 9.

[8] OECD DAC (2012) *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility: Improving Learning for Results*. Paris: OECD.

[9] DfID (2013) *International Development Evaluation Policy*. London: DFID. Para. 54. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/204119/DFID-Evaluation-Policy-2013.pdf; OECD DAC (2002) *Glossary of Key Terms in Evaluation and Results-Based Management*. Paris: OECD. p. 24.

[10] DfID (2012) *How to note: Results in Fragile and Conflict-Affected States and Situations*. Para. 9.16.

In addition, there is an older use of the term impact, which adds an additional layer of potential confusion, and is often interchangeable with the criterion of 'sustainability'. In this use of the term, impact is understood to mean a change that is long-term, sustainable, and can be seen as a lasting legacy of an intervention.

The fact that there are multiple conceptions of 'impact' may not be well understood, and the varied use of the term can create further confusion when combined with the term evaluation. Three different interpretations of 'impact evaluation' exist of interest for evaluation of peacebuilding that emphasise different aspects of the notion of impact.

1. *Attribution of specific outcomes to an intervention.* In this interpretation, impact evaluation is understood to mean assessment of the net effect of the intervention, as determined by what would have happened in the absence of the programme. It tends to emphasise quantifying the effect. It focuses on establishing cause and effect and 'must include a counterfactual or similar analysis to address attribution and establish causality'[11]; this is the interpretation held by many professional evaluators; it does not specify what kinds of effects are to be investigated[12].

2. *Effects of interventions on drivers of conflict.* In this interpretation, impact evaluation is understood to mean assessment of the intervention on the higher-level changes in the conflict situation to which it aims to contribute. The OECD DAC guidance, *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility*, notes that '[i]n 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'[13].

3. *Sustained outcomes of an intervention.* In this interpretation, impact evaluation in peacebuilding is defined as seeking 'to determine the change in the conflict catalysed by a project… [impact evaluations] are implemented at points ranging from several months to several years after the project is finished'[14]. This definition integrates the OECD DAC definitions of impact (described above) and of sustainability and underscores that some long-term effects are only witnessed after many years.

These three interpretations of 'impact evaluation' are not mutually exclusive, but can create considerable confusion. As Howard White, executive director of the International Initiative on Impact Evaluation (3ie) has noted, 'These are completely different definitions of impact', and 'neither side is

---

[11] *DfID Evaluation Policy*, para. 77, 79.

[12] See Leeuw, F. & Vaessen, J. (2009) *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation*. Washington, D.C.: Network of Networks on Impact Evaluation; Rogers, M. (2012) *Evaluating Impact in Conflict Prevention and Peacebuilding Programs*. Cambridge, MA: CDA. http://cdacollaborative.org/media/123704/Evaluating-Impact-in-Conflict-Prevention-and-Peacebuilding-Programs.pdf. *See also* the International Initiative for Impact Evaluation (3ie), which defines impact evaluation as and evaluation that "measure[s] the net change in outcomes amongst a particular group, or groups, of people that can be *attributed* to a specific program using the best methodology available, feasible and appropriate to the evaluation question(s) being investigated and to the specific context". 3ie (no date) Principles for Impact Evaluation. http://www.3ieimpact.org/media/filer_public/2012/05/17/principles_for_impact_evaluation.pdf.

[13] OECD DAC (2012) *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility, p. 67.*

[14] Church, C. & Rogers, M. (2006) *Designing for Results: Integrating Monitoring and Evaluation in Conflict Transformation Programs*. Washington, D.C.: Search for Common Ground. http://www.sfcg.org/programmes/ilt/ilt_manualpage.html.

right or wrong'[15]. They reflect different degrees of concerns with two central concerns of evaluations of impact: what are the nature and level of impacts we are concerned about? How do we assess the intervention's role in producing them?

Understanding causation is at the heart of the term 'impact evaluation', whichever interpretation of 'impact' is focused on. Any evaluation of peacebuilding impact will need to investigate whether the intervention was a cause of (or contributed to) observed impacts. Causation seeks to connect the intervention with the effects, attributing if, how, and how much a given intervention 'caused' a particular effect.

For those commissioning evaluations of impact, it is crucial to define and negotiate agreement about what is meant by impact and impact evaluation with country offices, implementing partners, and other key stakeholders before drafting Terms of Reference (ToRs) and evaluation proposals, as this will prevent misunderstandings and strongly inform the shape, methods and use of the evaluation.

2.2   Why conduct an impact evaluation?

Whichever interpretation is emphasised, impact evaluations can be valuable in generating rigorous evidence and learning about what works and why, and, when findings are used, in ensuring that lessons are learned during interventions and resources allocated to where they can have most effect. If attribution is the focus, an impact evaluation can provide valuable evidence about whether an intervention actually led to the results observed. For more macro level effects, they can help practitioners and policy makers assess whether their strategies are relevant and valid, and whether and how their assistance works with other interventions to ameliorate a conflict situation.

However, not all interventions merit an impact evaluation. Impact evaluations require time, resources (both financial and human) and often considerable technical expertise. They are not an add-on to a different evaluation process that can be conducted in a matter of hours or simply based on the evaluator's expertise. It is therefore useful to assess carefully in advance whether evaluating impact would be useful.

Consideration should be given to whether this evidence is *needed* at a particular point in time, and whether the impacts of the intervention are in fact evaluable. To justify the technical and financial resources necessary to carry out a high-quality evaluation of impacts, the intervention to be evaluated should meet at least one of the following criteria:

- *Significant enough* The programme or policy should be significant enough (in terms of size, policy importance, or potential consequences) to merit an evaluation of impacts.
- *Strategically relevant* to conflict and peace.
- *Untested* in the sense that there is not a great deal of evidence, or there is contested evidence or differing views about whether this kind of intervention is effective in the context in which it is being undertaken.
- *Influential* The results of the evaluation will be used to inform key decisions about policy and programming
- *Evaluable* For example data will available and outcomes evaluable; the timing is early enough to allow for observation at different points in time and to facilitate people's recall, but late enough so

---

[15] White, H. (2009) *Some Reflections on Current Debates in Impact Evaluation*. 3ie Working Paper 1. Washington, D.C.: 3ie. p. 6.

that outcomes and results can be seen; there is sufficient budget for a robust evaluation, and the answers to the evaluation question(s) will justify the cost[16].

For evaluations of this significance and cost, an evaluability assessment may be advisable. If the case for impact evaluation cannot be made, or is weak, different evaluation approaches relating to relevance, outcomes, implementation and sustainability may be in order. Several useful resources on evaluability assessments, both in general and tailored to peacebuilding interventions, exist:

> Davies, R. (2013) *Planning Evaluability Assessments: A Synthesis of the Literature with Recommendations*. Report of a Study Commissioned by the Department for International Development. London: DfID. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/248656/wp40-planning-eval-assessments.pdf.

> Reimann, C. (2013) *Evaluability Assessments in Peacebuilding Programming*. CDA Working Paper on Program Review & Evaluation #3. Cambridge, MA: CDA.

> Rogers, M. (2013) *Evaluating Impact in Conflict Prevention and Peacebuilding Programs*. CDA Working Paper on Program Review & Evaluation #2 (Section II). Cambridge, MA: CDA.

2.3   Challenges of evaluating impact in peacebuilding interventions

In addition to the confusion about what 'impact evaluation' means, the nature of peacebuilding interventions and conflict contexts complicate peacebuilding evaluations. A number of factors make it difficult to identify and agree on what should be analysed or evaluated, how to evaluate, and how to infer with rigour any causal links between the intervention and changes in the conflict context[17].

- *Defining 'success' is difficult and contested.* Peacebuilding has been defined in many ways—stabilisation and reduction of violence, building a 'just and sustainable' peace addressing underlying grievances and injustices, and statebuilding[18]. The types of interventions are wide-ranging—and their intended objectives often vague, at a very high level, or constantly changing in response to the evolution of the conflict situation. The diverse, and often debated, notions of what 'peacebuilding' is make it difficult to define relevant measures by which to evaluate. Some frameworks, such as the Peace and Conflict Assessment (PCIA) and the Measuring Progress in Conflict Environments (MPICE), set out more concrete categories and indicators of peacebuilding impacts that should be evaluated[19]. These can be useful in focusing inquiry, but risk creating criteria for judging the impact of an intervention that are based on a normative definition of peacebuilding, and can result in assessing progress against a hoped for but largely impossible end-state. In other words the criteria

---

[16] Adapted from Gertler, P., Martinez, S., Premand, P., Rawlings, L. & Vermeersch, C. (2011) *Impact Evaluation in Practice*. Washington, D.C.: World Bank, p. 10; and Rogers, M. (2012) *Evaluating Impact in Conflict Prevention and Peacebuilding Programs*. CDA Working Paper on Program Review & Evaluation #2. Cambridge, MA: CDA.

[17] See Andersen, O. & Kennedy-Chouane, M. (2014) 'Introduction'. In Andersen, O., Bull, B. & Kennedy-Chouane, M. *Evaluation Methodologies for Aid in Conflict*. London: Routledge.

[18] *Id. See also* Anderson, M. & Olson, L. (2003) *Confronting War: Critical Lessons for Peace Practitioners*. Cambridge, MA: CDA.

[19] Bush, K. (2005) *Hands On PCIA: A Handbook for Peace and Conflict Impact Assessment (PCIA)*. Ottawa: St Paul University. http://www.peaceprize.uclg.org/fileadmin/user_upload/HOP--_Handbook__--_2008_June.pdf; Agoglia, J., Dziedzic, M. & Sotirin, B. (2010) *Measuring Progress in Conflict Environments (MPICE): A Metrics Framework*. Washington, D.C: USIP. http://www.usip.org/publications/measuring-progress-in-conflict-environments-mpice. The Peace and Conflict Impact Assessment (PCIA) lists five areas for investigation: conflict management capacities, militarized violence and human security, political structures and processes, economic structures and processes and social empowerment.

reflect the evaluator's preferred theory, rather than using context-appropriate and locally-accepted criteria that reflect what is happening on the ground[20].

- *Examining the causes of indirect effects is difficult.* It is much easier to make causal linkages in short results chains—i.e. when the number of steps or changes between the activities implemented and the impacts is small. The diverse types of activities conducted for peacebuilding—from dialogue and reconciliation programmes to reintegration of ex-combatants, justice reform, infrastructure development and livelihoods—do not all work directly on conflict and peace, but rather seek to affect conflict and peace drivers indirectly. Evaluating this impact entails examining an intervention not only in terms of its own objectives, but (also) in terms of actual outcomes as related to key drivers of conflict and peace. This creates longer and more complex causal chains, and makes it impossible to make defensible judgements about sole attribution.

- *Evaluating broader peacebuilding impacts of projects and programmes is difficult.* The OECD DAC's guidance on *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility* provides examples of questions to examine the criterion of impact, including: 'what key drivers of conflict and fragility were affected and how?' 'Is the programme contributing to changes at the communal or societal level (beyond participants)?' Answering these questions is difficult at the project or programme level, as the contribution of any single programme to changes in society is likely to be quite small and hard to observe. As result, some analysts believe that impact should only be assessed at the strategic or portfolio level, or when a single intervention is high-level (e.g. high-level mediation effort). At the same time, evaluation of peacebuilding impacts only at the strategic level can relieve individual agencies of responsibility for thinking rigorously about how they contribute to broader impacts, and can limit the usefulness of evaluation results for them because the findings are about cumulative effects.

- *Complexity makes determining effects difficult.* There are many variables that influence how change comes about in complex and non-linear ways. This means that changes are likely to take place at different speeds in different arenas, because of a combination of different efforts interacting with contextual factors. It is difficult to isolate an intervention's impacts from the many other factors and to know whether impact will be observable at the time of an evaluation. Some interventions, for example, may prepare the ground for changes that occur only after later events or interventions trigger them (e.g., a negative event that galvanises people to action, a mediation process, a land reform initiative).

- *Peacebuilding results or contributions may not be visible.* Many peacebuilding activities are confidential, politically sensitive or rely on participants fully adopting them as their own. The success of this work is achieved without others being aware of it, and this success occurs *because* no one knows. Bringing it out in the open through impact evaluation can undermine effects and undermine participant ownership.

- *Political nature of peacebuilding.* Peacebuilding work is often highly political in nature. Issues addressed in peacebuilding—such as restructuring of the security sector, land reform, power-sharing, corruption, or transitional justice, among others—are politically sensitive. Moreover, aid provided in conflict-affected or fragile environments is often part of broader geopolitical or other agendas. This means that interventions may have vague, hidden, ambiguous or overly ambitious goals that make identifying what 'impacts' are to be evaluated difficult. Even when they can be

---

[20] Stave, S. (2011) *Measuring peacebuilding: challenges, tools, actions*. NOREF Policy Brief No. 2. Oslo: NOREF. http://www.peacebuilding.no/var/ezflow_site/storage/original/application/906762cb32e2eed5dc810bafa139f4ce.pdf.

evaluated, the process and the sharing of the results take place in a political (and sometimes politicised) environment and can be distorted and used in ways that may undermine the intervention and its effects.

3 Key Design Considerations for Evaluating Impact of Peacebuilding Interventions

Evaluation approaches refer to the principles or framework guiding the design and implementation of the evaluation. There has been considerable debate about which approaches and methods are best for impact evaluation. While some argue that experimental and quasi-experimental methods should be used wherever possible to prove causal links[21], others claim that more qualitative, process-based approaches 'often lead to stronger causal influences than statistical or experimental designs'[22].

Situational appropriateness is increasingly being seen as the best criterion for choosing methods. Commissioners of evaluations, together with evaluators, should decide what methods are appropriate based on:

- the purpose of the evaluation
- the users' needs and intended uses of the evaluation
- the nature of the intervention (e.g., is it a standardised intervention, such as an infrastructure intervention, or an adaptive one that needs to change and adapt to the context?)
- the availability of resources
- the nature of the impacts sought (e.g. produced directly by the intervention, or indirectly, like a ripple; short-term or visible only after the programme has ended, immediate and within the programme's sphere of influence or macro, etc.) and
- whether the intervention is affected by and interacts with contextual factors to generate the desired impacts[23].

Box 1: Fitting the evaluation design to the intervention and the setting

'The idea that there is a single superior method of producing evidence has also been widely challenged…Rather it is generally understood that methods and designs are fit for different purposes and when well-executed all have their strengths and weaknesses—one of the justifications of so called "mixed-methods". Furthermore the choice of methods (and overall designs) needs to follow from the kinds of questions that are being asked; whilst also taking into account the settings in which they are to be applied. This underlines the importance of identifying a range of designs; and understanding when each is appropriate.'

Source: Stern, E. et al. (2012) *Broadening the range of methods and designs for impact evaluation—report of a study commissioned by the Department for International Development*. DfID Working Paper 38. Para 2.21.

Many approaches and methods used for development evaluations are also useful in peacebuilding situations—with some adaptations. Following are a number of special considerations to be taken into

---

[21] *See* Gaarder, M. & Annan J. (2014) 'Impact evaluation for peacebuilding: Challenging preconception'. In Andersen, O. et al. *Evaluation Methodologies for Aid in Conflict*. Chapter 7.

[22] Khagram, S., Thomas, C., Lucero, C. & Mathes, S. (2009) 'Evidence for development effectiveness.' *Journal of Development Effectiveness*. Vol. 1, No. 3, p. 258.

[23] See Rogers, P. (2009) 'Matching impact evaluation design to the nature of the intervention and the purpose of the evaluation'. *Journal of Development Effectiveness*. Vol. 1, No. 3, pp. 217-226.

account in deciding what design and what methods might be most appropriate for impact evaluations of peacebuilding interventions.

3.1 Evaluation focus—what is the nature and level of impacts to be evaluated and what questions can an evaluation of peacebuilding impact focus on?

3.1.1 Level of impacts

The type of intervention and the scale or level of impact(s) sought are important determinants of what evaluation focus and design might be appropriate. A peacebuilding intervention may affect multiple individuals, groups, institutions or broader issues. For example for a programme that supports local inter-group economic initiatives, the intended impacts may include: improved livelihoods opportunities for populations at risk of conflict, improved cross-group economic and social relationships, increased perceptions that cooperation is more beneficial than conflict, resistance to violence or provocations to violence, greater ability (e.g., skills, mechanisms) to resolve community conflicts and reduction of violence[24]. These impacts are different, and operate at different levels; it is important to understand what level it is useful to focus the inquiry at, given the purpose of the evaluation. Issues of causality—whether and how the intervention produced or contributed to the results or 'impacts'—are present at all levels; however, they would be addressed differently, and different methodological options would be available and appropriate at each level.

What level and type of impact should we be interested in? There are many possible levels at which an impact evaluation might be focused. These implicitly underlie the three interpretations of 'impact evaluation' (see Section 2), but should be considered explicitly:

- *Impacts of the programme on its intended participants or partners.* The evaluation may want to look at whether the programme produced the direct effects listed in its objectives. For example in a programme that supports local inter-group economic initiatives, did it increase wealth or livelihoods opportunities for participants? Did it lead to inter-ethnic economic cooperation? Did participants perceive tangible benefits from cooperation, and did they believe conflict would threaten those? Here the evaluation might focus on whether any changes were *attributable* to the programme, as opposed to other factors.

- *Impacts on drivers of conflict.* The evaluation may want to look at broader indirect or second-order impacts—such as lessening of local grievances, reduction of incentives to participate in violence, or reduction in economic inequalities among conflicting groups[25]. In the inter-ethnic cooperation programme above, did inter-ethnic cooperation expand beyond the participants? Did inter-ethnic trust and relationships improve? Were communities able to resist provocations to violence? Did inter-group disputes over land decrease? This entails looking beyond the immediate scope of the programme to its wider effects, identifying any changes that have occurred in conflict and peace drivers over the relevant time period (in the location, regionally or nationally) and assessing whether and how changes for the participants have transferred beyond the participants to the broader context and influenced the drivers and actors of peace and fragility.

---

[24] *See* Mercy Corps (2010) *Evaluation and Assessment Poverty and Conflict Interventions: Conflict & Economics: Lessons Learned on Measuring Impact*. Portland, OR: Mercy Corps; Mercy Corps (2011) *Building Bridges to Peace: Final Evaluation Report.* Washington, D.C.: USAID.

[25] The development assistance-focused NONIE guidance distinguishes between similar 'levels' of impacts: beneficiary level and institutional level impacts—the latter representing second-order conditions such as capacities, willingness or structures enabling institutions to develop and implement better policies for communities, households and individuals. Leeuw, F. & Vaessen, J. (2009) *NONIE Guidance*, pp. 31-32.

- *Long-term and sustained impacts.* When a programme is completed, summative evaluations often utilise the OECD DAC criterion of *sustainability* to analyse 'the continuation benefits on end of assistance', including continued 'resilience to risk' over time and ownership of the peace and development processes[26]. This assesses the *probability* of continuous change by determining the levels of local or national ownership of short-term outcomes; and the likely durability of new or existing institutions or mechanisms. To understand whether a programme has had sustained long-term impact, evaluators must return to the geographic location where the project was implemented to evaluate whether the intermediate outcomes or changes produced during the life of a project have been sustained and/or continued; whether they ultimately shifted conflict drivers; and/or whether they created a culture that can effectively mitigate conflict through peaceful means. For example in the example above did economic cooperation continue after the programme ended? Did the perceptions of benefits of cooperation continue? Did changes in attitudes, governance processes, or changes in wealth continue after the end of the programme? Did participants resist violence or provocations to violence long after the programme ended?

Box 2: The need to evaluate sustained impacts

Evaluations of sustained impacts can provide valuable information about how and what transformational effects a peacebuilding intervention can have. An impact evaluation of a dialogue programme bringing students from the United States together virtually with students from Arab countries focused on inter-group attitudes and affect. The evaluation assessed attitudes and feelings toward the other group (e.g., US about Muslims) among programme participants and a control group. At the end of the programme, it found significant effects on feelings toward the other, and some on attitudes about whether there was a 'clash of civilisations' as compared to the control group. The evaluator administered a survey four months later, finding that effects on feelings had faded, while effects on cognitive attitudes (among Americans) continued to increase. While this is not an evaluation of 'sustained impacts', the findings underline the utility and importance of returning to the site of the programme or the participants to understand whether (and what) impacts have been sustained, to generate learning for future peacebuilding programming.

Source: Authors' conversation with evaluator

In peacebuilding interventions, impacts at the participant-level are often stepping stones toward effects on the broader peace, or 'peace writ large', and considerations of whether the changes can be sustained and owned once the intervention is completed are critical for understanding whether progress is being made and whether real transformation of the conflict is taking place.

3.1.2   Evaluation questions

Situational appropriateness involves matching the design of any evaluation to the desired purpose, the level of impacts being assessed, and the constraints and opportunities of the particular context. Clear evaluation questions tailored to the evaluation's purpose and the nature and level of the intervention's impacts help frame what methodologies will be appropriate, feasible and useful and how causal relationships between the intervention and the 'impacts' of interest can be assessed.

For those commissioning an impact evaluation, the most important place to start is with the purpose and uses of the evaluation. These questions help decide whether an impact evaluation is appropriate, and if so what should be evaluated, what questions need to be asked in the evaluation and what kind of evaluation design and methods that will most suit the purpose, and are likely to help the users of evaluation. For example:

- Is a pilot or innovative approach being considered for scale-up or replication?

---

[26] OECD DAC (2012) *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility,* p. 68.

- Is there a question about the efficacy of possible alternative interventions to achieve the same goal (e.g., two different disarmament packages or two different approaches to reintegration) which an impact evaluation comparing them might help answer?
- Is the intervention significant—for example because of its scope or size, initial investment, political importance, importance of the issues it addresses or outcomes that it wants to affect?
- Is there disagreement or inconsistent evidence about whether a peacebuilding strategy, theory of change or programme model works in any given context?
- Is a decision being made about whether to continue a strategy or programme for which evidence of whether it supporting peace would be helpful?

The purpose of the evaluation, along the level of impacts being assessed and the constraints and opportunities of the particular situation should be the basis for determining evaluation questions. The table below provides examples of key evaluation questions for impact evaluation of peacebuilding interventions. It is better to focus on a small number of questions directly related to the purpose of the evaluation, rather than spread resources and users' focus across a large number of questions that may be interesting, but not critical.

Table 1: Illustrative evaluation questions for impact evaluations of peacebuilding interventions

Examples of key evaluation questions for peacebuilding	
Common questions at all levels	
Overall impact: • Did the intervention work? Did it produce the intended impacts in the short, medium and long terms? • What unintended impacts (positive or negative) did the intervention have on drivers of conflict and fragility or on dividers and connectors in the context? • To what extent is the theory of change valid?	**How and why the intervention works:** • How did the intervention contribute to the intended impacts, maintenance of impacts over time, or wider or macro-level impacts on conflict? • What features of the intervention made a difference? • Are differences in impact explained by differences in implementation? • To what extent are failures or lack of impact due to implementation vs. theory of change? • How and why were impacts sustained over the long term?
Influence of other factors: • How did the intervention work in conjunction with other interventions or activities to achieve impacts? • What features of the context of conflict and fragility enhance or limit wider impacts of the intervention?	**Replication:** • Can this intervention (model, theory of change) be expected to work elsewhere? How and under what circumstances?
Programme/Participant-level impacts	*Wider/impacts on conflict drivers and sustained impacts*
• To what extent can the specific impacts be attributed to the intervention? • Did the intervention make a difference in relationships, attitudes, behaviours and structures relevant to conflict and peace drivers?	• Did the intervention contribute to changes at the communal or societal level (beyond participants)? • Did the intervention contribute to changes in key drivers of conflict and fragility? • How and why were impacts sustained over the long term?

Source: Adapted from Rogers, P. (2012) *Introduction to Impact Evaluation*. Impact Evaluation Notes, No. 1. Washington, D.C.: Interaction and Stern, E. et al. (2012) *Broadening the Range of Designs and Methods for Impact Evaluations*. DfID Working Paper 38. London: DfID.

3.2   Conducting conflict analysis and analysis of relevance

> The impact of development cooperation can only be assessed against clearly defined relevance criteria. In conflict zones, these criteria need to be developed by analysing the conflict and identifying main threats to peace…Conflict zones are characterized by volatility, and not everything in the context is equally relevant to understanding the conflict. A single focus on "root causes" of a conflict or a long list of factors potentially affecting stability can distort an understanding of the actual dynamic of the conflict in which development cooperation takes place[27].

A distinguishing characteristic of evaluations of impacts in peacebuilding is that the changes an intervention has produced or contributed to must relate to the dynamics or drivers of conflict and peace. It is not enough to demonstrate that the intervention has contributed to changes in the *context or the individual*; the intervention must be relevant to those aspects of the context of conflict or fragility that directly shape or influence how the conflict develops and whether violence occurs[28]. If not, then the intervention may do good and have impacts, but still not change any underlying dynamics or drivers of conflict, and therefore not have *peacebuilding* impacts.

3.2.1 Why conflict analysis and relevance analysis?
Conflict analysis is therefore key to evaluation of impacts of peacebuilding programming and policies, along with an analysis of the ***relevance*** of the intervention to the conflict. Conflict analysis helps to understand key driving factors of conflict and peace and the evolution of those drivers over the time period covered. Relevance analysis identifies whether the broader changes targeted by the intervention are connected to the needed changes specified in the conflict analysis and thus whether the intervention makes a logical contribution to meeting peacebuilding needs. Together these types of analysis help evaluators a) identify any change in the conflict situation to which the programme intended to contribute, and b) identify and unpack the hypotheses or theories of change of the intervention in relation to peacebuilding—i.e. the assumptions of how it will contribute to addressing the drivers of conflict—so they can be tested in the evaluation.

3.2.2. Identifying conflict drivers and relevance at different levels of impact
For an impact evaluation looking at the ***wider effects*** of an intervention on macro-level conflict drivers, the conflict and relevance analysis provides the benchmark against which the intervention will be assessed, i.e. whether it contributed to change in the key drivers of the conflict. This is not difficult for evaluations where the scale or significance of the intervention makes it reasonable to expect a direct effect at the macro level. However, for individual programmes or smaller programmes working at community level, this is more difficult, as their effects at the macro level are likely to be only indirect and achieved in combination with many other interventions and contextual factors.

---

[27] Bohnke, J., Koehler, J. & Zurcher, C. (2014) 'Assessing development cooperation in northeast Afghanistan'. In Andersen, O. et al. *Evaluation Methodologies for Aid in Conflict*. p. 107.
[28] *See* OECD DAC (2012) *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility*, p. 45.

When wider impacts of an intervention cannot be observed at the macro level (e.g., for many smaller or community-level individual programmes or projects) the conflict analysis can be used to identify mid-level effects or 'enabling conditions' that reflect macro-level changes but at a level or scale appropriate for the intervention. In its evaluation of DfID's Conflict Pool, for example, ICAI disaggregated the overall ('peace writ large') goal of improving India-Pakistan relations into a number of objectives (such as sustained dialogue between the countries, reduced violence in Kashmir, changed attitudes in both countries) and identified 'enabling conditions' necessary to achieve those objectives. Evidence of measurable progress in achieving the enabling conditions could provide a credible story of impact[29].

CDA's Reflecting on Peace Practice program has also identified several criteria for identifying impacts on 'peace writ large' at the programme or project level that can be useful for identifying such mid-level results.

Table 2: Illustrative criteria and lines of inquiry for projects and programmes

Reflecting on Peace Practice (RPP) 'Building Blocks of Peace'
From analysis of case studies of peacebuilding practice and practitioner reflection on their own experiences, the RPP process identified five intermediate Building Blocks that can support progress towards Peace Writ Large. These can be used to assess, across a broad range of contexts and programming approaches, whether a program is making a meaningful contribution to 'peace writ large'. The Building Blocks can be used in program planning and evaluation to ensure that specific program goals are linked to the larger and long-term goal of contributing to 'Peace Writ Large'.
Does the intervention address key dynamics and drivers of conflict and peace AND: 1. Are institutions or mechanisms (formal or informal) that address the specific grievances or injustices that fuel the conflict created or reformed? 2. Does the effort causes participants and communities to develop independent initiatives that decrease

---

[29] Independent Commission for Aid Impact (ICAI) (2012) *Evaluation of the Inter-Departmental Conflict Pool*. London: ICAI.

dividers, increase connectors or address causes of conflict?

3. Does the effort prompt people increasingly to resist violence and provocations to violence?
4. Does the effort result in an increase in people's security and in their sense of security?
5. Does the effort result in meaningful improvements in inter-group (collective) relationships?

Finally, where the evaluation seeks to assess attributable impacts on participants or beneficiaries of the intervention within its sphere of influence, conflict analysis and analysis of relevance is needed to ensure the 'impacts' to be measured are actually relevant to conflict and peace drivers. For example, if land issues are significant drivers of macro-level conflict, then a conflict resolution mechanism or community council might be measured on its ability to address land conflict.

3.2.3 Performing the conflict analysis

Ideally, a conflict analysis will have been conducted at the outset of a programme, or as a country operational plan or a programme is being developed and monitored, so that evaluators would not need to conduct an analysis themselves—but merely verify one that exists. The conflict analysis should provide the basis for the identification of key indicators and issues to be measured in a baseline and monitored and updated throughout the life of the programme or operational plan. It also provides a point of comparison for the endline, permitting identification of any changes that may have occurred over the course of the programme or operational plan.

In many cases, however, such systematisation and monitoring of the conflict has not been done, and the evaluation must reconstruct the analysis to be used as the criteria for assessing impact. Different approaches can be used to reconstruct conflict analyses. For example, combining desk review, with localised participatory analyses of events and trends involving stakeholders of various backgrounds in different localities[30], or using local newspapers to create a dataset of reported conflicts in the research areas and surrounding districts[31]. In the multi-donor evaluation of conflict prevention and peacebuilding activities in Southern Sudan, the team conducted two analyses: one aiming to capture what donors knew or should have known in 2005, and the other an updated (2010) analysis of the conflict factors and the extent to which donors had (or had not) adapted to the changing situation over time[32].

3.3   Importance of theory of change and theory-based evaluation approaches

The OECD DAC guidelines on *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility* advise identifying and assessing the theory of change of an intervention[33]. This is a critical step, in understanding and testing the assumed mechanisms for change that underpin the intervention.

Theories of change are used in two ways in evaluations of peacebuilding impact. First, they provide the basis for constructing the 'causal story' of the intervention—how and why it produced results—and thus for establishing attribution and contribution. The causal story can then be tested through a variety of methods. Second, theories of change are used as an explicit benchmark for testing some of the assumptions in a more formal manner. Each 'link' in the causal chain can provide a hypothesis that could be tested formally and experimentally.

---

[30] Brusset, E. et al. (2011) *Amani Labda (Peace Maybe): Joint Evaluation of Conflict Prevention and Peacebuilding in the Democratic Republic of Congo*.

[31] Barron, P. et al. (2006) *Local Conflict and Community Development in Indonesia: Assessing the Impact of the Kecamatan Development Program*, Indonesian Social Development Paper No. 10. Jakarta: World Bank.

[32] Bennett, J. et al. (2010) *Aiding the Peace: A Multi-donor Evaluation of Support to Conflict Prevention and Peacebuilding Activities in Southern Sudan 2005-2010*. United Kingdom: ITAD, Ltd.

[33] OECD DAC (2012) *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility*, p. 58.

This guidance does not address how to develop or articulate a theory of change. Information and guidance on theories of change, how to develop and articulate them at different levels and how to use them in evaluations is covered in other Guidance Products of the CCVRI and are relevant to impact evaluations:

> Woodrow, P. & Oatley, N. (2013) *Theories of Change in Conflict, Security & Justice Programmes, Part I: What they are, different types and how to develop and use them*.

> Corlazzoli, V. & White, J. (2013) *Theories of Change in Conflict, Security & Justice Programmes, Part II: Using Theories of Change in Monitoring and Evaluation*.

In impact evaluations, identifying or reconstructing theories of change in evaluations of peacebuilding interventions can be challenging for a number of reasons:

- Peacebuilding theories are often incomplete, or articulated vaguely with big gaps in logic.
- Peacebuilding objectives tend to be poorly written, often vague, blurry or overambitious.
- External forces related to the conflict or peace are strong and difficult to manage, and theories of change evolve as interventions adapt to deal with changing circumstances.
- Peacebuilding interventions often use development, humanitarian, human rights, security and other modalities whose immediate goals and objectives are not related to conflict; the assumptions about how those goals relate to and affect conflict are often implicit, or very general—for example '"a contribution towards establishing the peace" or a "peace dividend"'[34].
- In multi-faceted initiatives, the same activities may be undertaken based on different theories of change; these may be compatible, but may be competing, and articulating a single 'theory of change' is impossible.
- Documentation and monitoring of the theory of change (and of outcomes) throughout the life of the intervention is poor.
- High staff turnover in fragile and conflict-affected contexts often leads to loss of institutional memory; staff (both national and international) involved in interventions at the time of an evaluation are often not those who were involved in the design, and in the case of evaluations of sustained impacts, have often dispersed and are difficult to find.

---

[34] Barnett, C. & Bennett, J. (2014) 'Critical reflections on the South Sudan evaluation of conflict prevention and peacebuilding activities'. In Andersen, O. et al. *Evaluation Methodologies for Aid in Conflict*. p. 52.

When theories of change are vague, incomplete or hard to reconstruct, evaluators have taken a variety of approaches to overcome this challenge. The *Thematic Evaluation of European Commission Support to Conflict Prevention and Peacebuilding* (2011), for example, which assessed the Commission's 'integrated approach' to peacebuilding, reconstructed the programme logic representing the theory of change based on official policy documents and mapped the evaluation questions on to it[35]. Other evaluations have used conflict analysis, in conjunction with portfolio reviews, reviews of the activities and outcomes of programmes, stakeholder interviews and research to construct hypotheses about how the intervention influences conflict drivers (see Box 5 for an example). Such hypotheses are similar to theories of change but do not entail reconstructing a hierarchy of results as *intended* by any agency; they also have the advantage of being able to synthesise differing assumptions and perspectives of stakeholders into an overarching theory. Where the assumptions or expectations are so different that they cannot be synthesised, it is possible to elaborate competing theories that can be tested.

Box 5: Developing 'working hypotheses' instead of 'theories of change'

Faced with challenges in reconstructing theories of change, the multi-donor evaluation of conflict prevention and peacebuilding in southern Sudan, which focused on evaluating the overall impacts of donor assistance, developed 'working hypotheses' of how the interventions were expected or assumed to influence conflict factors. '[T]he evaluation's conflict analysis informed both the selection of the "working hypotheses" for testing in the field, and led to findings that would not necessarily have emerged when evaluating an intervention within its own terms (and objectives). For example, the evaluation found that 'the confusion between "marginalisation" and 'lack of development' led to an assumption that lack of development was not simply a matter of concern but a factor causing conflict', opening the way 'for what became the dominant "theory of change": that all forms of development contribute' to peacebuilding. This hypothesis was tested in the evaluation and found to be erroneous; 'in Southern Sudan there was no evidence that conflicts might arise from…lack of services'.

Sources: Barnett, C. & Bennett, J. (2014) 'Critical reflections on the South Sudan evaluation of conflict prevention and peacebuilding activities,' in Andersen, O. et al., *Evaluation Methodologies for Aid in Conflict*. London: Routledge, pp. 39, 49; Bennett, J., Pantuliano, S., Fenton, W., Vaux, A., Barnett, C. & Brusset, E. (2010) *Aiding the Peace: A Multi-donor Evaluation of Support to Conflict Prevention and Peacebuilding Activities in Southern Sudan 2005-*2010. United Kingdom: ITAD Ltd, p. 49.

3.4   Combining 'inside out' and 'outside in' analysis

Combining conflict analysis with an understanding of the theories or assumptions about how interventions are supposed to address conflict drivers allows impact evaluations to provide both 'inside out' and 'outside in' analysis[36]. The 'outside in' analysis uses context and conflict analysis to gain an understanding of how a conflict has evolved, as well as what the constraints and opportunities for an intervention's influence on it are at any particular moment in time. Some evaluations, such as that of USAID's 'People-to-People Reconciliation Annual Fund', specifically request assessment of larger contextual developments that restrict or facilitate wider impacts of programming[37]. The 'inside out'

---

[35] ADE (2011) *Thematic Evaluation of European Commission Support to Conflict Prevention and Peacebuilding: Final Report*. Louvain: ADE.
http://ec.europa.eu/dgs/fpi/documents/thematic_evaluation_of_ec_support_to_pb_and_conflict_prevention_2011_en.pdf.

[36] These terms were used by the evaluators of Norway's facilitation of the Sri Lankan peace process to describe their overall approach to the evaluation. *See* Goodhand, J., Klem, B. & Sorbo, G. (2014) 'Battlefields of Method: Evaluating Norwegian peace efforts in Sri Lanka'. In Andersen, O. et al. *Evaluation Methodologies for Aid in Conflict.* Chapter 4.

[37] Allen, S. et al. (2014) *Evaluative Learning Review Synthesis Report: USAID/CMM's People-to-People Reconciliation Fund, Annual Program Statement (APS)*. Washington, D.C.: Social Impact. Available at http://dmeforpeace.org/sites/default/files/CMM%20Evaluative%20Learning%20Review_Synthesis%20Report%20

approach looks at the intervention itself, its activities, outputs and outcomes and the extent to which these are reflected in the working (rather than assumed) theory of change. By linking these two perspectives, evaluators can minimise issues of bias by placing the intervention in the larger context in which it is operating (and of which it is a part) and can identify the range of factors and dynamics, including the intervention, that have influenced the evolution of the conflict.

These types of analysis are particularly important in evaluations focusing on wider impacts of peacebuilding (beyond participants, on drivers of conflict) and long-term sustained impacts. This is where constraints and opportunities, as well as effects catalysed by an intervention in combination with other factors, have significant effects on the nature, scale and scope of 'impacts'. It also reflects a growing recognition both among researchers and practitioners in the development and peacebuilding fields that interventions are being carried out in systems that are too interconnected and too complex for linear models of cause and effect to be useful[38].

Box 6: Combining 'inside out' and 'outside in' approaches for impact evaluation in peacebuilding: Examples from Sri Lanka and Kosovo

The evaluation of **Norway's facilitation role in Sri Lanka** from 1999-2009 conducted an analysis of the broader structural context and conflict and peacemaking dynamics ('outside in') and developed a detailed account of Norway's involvement in the peace process ('inside out'). It linked the two by identifying key turning points in the situation in Sri Lanka and, using a process tracing methodology, scrutinising Norwegian responses in relation to what they knew and what the opportunities were at the time. This allowed exploration of the linkages between the mediator's decision-making and wider institutions, structures and dynamics.

In **Kosovo**, a study of the impacts of peacebuilding work on certain communities' success in avoiding violence during the widespread riots of 2004 combined 'outside in' and 'inside out' analysis in looking at cumulative effects of a variety of peacebuilding efforts. A broad conflict analysis of Kosovo was conducted and supplemented by in-depth analysis (quantitative and qualitative) of the nature and trends of violence across Kosovo over a four-year period. In addition, case studies of seven communities (selected for variation on density of peacebuilding activities, history of violence, and experience of violence in 2004) focused on the historical experiences and causes of violence and whether, how and why they were (or were not) able to avoid violence in the 2004 riots. The study also analysed peacebuilding programmes and identified broad theories of change underlying them. Because it did not focus on specific projects or programmes, the team identified types and density of peacebuilding programmes and their common assumptions about change. The two analyses were brought together to examine the interaction of programmes with drivers of violence and resistance/avoidance of violence and to assess the validity of the theories of change.

Sources: Goodhand, J., Klem, B. & Sorbo, G. (2011) *Pawns of Peace: Evaluation of Norwegian Peace Efforts in Sri Lanka, 1997-2009* (Oslo: NORAD evaluation 5/2011); Chigas, D. et al. (2006) *What Difference Has Peacebuilding Made? A Study of Peacebuilding and the March '04 Riots in Kosovo* (Prishtine/Pristina, Kosovo: CARE and CDA).

3.5   Conflict Sensitivity

The OECD DAC guidance on *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility* advises that 'Evaluations carried out before, during, or after a violent conflict must be conflict sensitive

---

Final%20March%202014_USAID_040714.pdf. The Review included field studies of people-to-people programming in West Bank/Gaza, Bosnia-Herzegovina and Burundi.
[38] See Ramalingam, B. (2013) *Aid on the Edge of Chaos: Rethinking International Cooperation in A Complex World*. Oxford: Oxford University Press ; Korppen, D., Ropers, N. & Giessman, H. Eds. (2011) *The Non-Linearity of Peace Processes—Theory and Practice of Systemic Conflict Transformation*. Opladen/Farmington Hills: Barbara Budrich Verlag; Chigas, D. & Woodrow, P. (2014) 'Systems thinking in peacebuilding evaluations: Applications in Ghana, Guinea-Bissau and Kosovo'. In Andersen, O. et al. *Evaluation Methodologies for Aid in Conflict*. Chapter 9.

because they are themselves interventions that may impact on the conflict…. Evaluators should keep in mind that the way they act, including both the explicit and implicit messages they transmit, may affect the degree of risk'[39]. Conflict sensitivity relates both to a) the effects of the conflict on the evaluation; and b) the effects of the evaluation on the conflict context.

Evaluation in fragile and conflict-affected contexts involves serious risks and serious real and immediate potential consequences for evaluators and for people participating in the evaluation. If evaluations are not properly conducted they may also increase the level of tension, worsen drivers of conflict or undermine cross-conflict connections and forces for peace. For example, small mistakes—such as an 'offensive' or inappropriate question on a survey—can escalate rapidly to bigger problems in conflict contexts. Technical issues, such hiring the wrong translator or driver, can quickly become conflict flashpoints—if, for example, the choice of translator is perceived as a preference for one conflict stakeholder over another[40].   Evaluators, as well as commissioners and managers of evaluations, thus need to be sensitive to and able to manage the interactions of the process with the conflict dynamics.

3.5.1   What are potential conflict sensitivity risks?
Conflict sensitivity problems often arise inadvertently from methodological choices and instruments, logistical challenges of conducting an evaluation in a conflict context, and political interests and efforts to influence an evaluation and its use. Risks include:

1. *Perceived (and real) partisanship and bias in evaluation process and results.* For example, a methodology that 'disappears' certain key stakeholders' voices because they are not included risks further marginalising already marginalised groups and escalating conflict if group inequality and exclusion are sources of grievance. This can occur when only voices representing one side's perspective on the conflict (e.g., one ethnic group, sect, clan, etc.) are included, either the methodology or evaluation process is 'conflict-blind', because of logistical and access issues, or because the methodology averages out results that are in fact different for different types of stakeholders. The choice of evaluation team members, even if clearly on technical merit, can also exacerbate tension and lead to accusations (and reality) of bias if they represent or are seen to represent any particular faction or perspective in the conflict. Bias can also result from the evaluator's personal, preconceived notions regarding the situation or the intervention that may distort data collection and reporting. Respondents in the evaluation may also be influenced by 'courtesy' and hospitality or because they want the programme to continue (or not) may not respond accurately. In peacebuilding interventions this is of particular concern as strong views about what causes conflict and what is needed for peace (e.g. 'liberal peace' theories) are often held by implementers and evaluators alike.

2. *Direct harm participants in the evaluation and to evaluators.* Security concerns are an omnipresent issue in evaluations in conflict zones, both for those participating in the evaluation and for evaluators—in some cases necessitating the cancellation of the evaluation itself. The mere presence of foreigners in some contexts, like Afghanistan, may subject them and evaluation respondents to security risks. However, harm can be caused by more indirect means. Evaluations can expose people to risk by directly or indirectly associating them with the subject of evaluation—for example if an

---

[39] OECD DAC (2012) *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility,* p. 35.

[40] The observations and frameworks for understanding conflict sensitivity of evaluation that follow in this section draw heavily on Colleen Duggan and Ken Bush's work on ethics of evaluation in conflict zones. See Bush, K. & Duggan, C. (2013) 'Evaluation in Conflict Zones: Methodological and Ethical Challenges'. *Journal of Peacebuilding and Development*. Vol. 8, No. 2, pp. 5-25; Duggan, C. & Bush, K. (2014) 'The Ethical Tipping Points of Evaluators in Conflict Zones'. *American Journal of Evaluation*. Vol. 35, No. 3, pp. 1-22.

evaluator visits someone in their home to interview them, or if sensitive information or potentially controversial views provided by respondents 'leak'. Questions posed in surveys and interviews can cause offense or lead to distress or reliving of traumatic events.

3. *Undermining of values of peacebuilding and development practice.* When participants in an evaluation are viewed as 'subjects' and approached (sometimes repeatedly) for information, and there is little transparency or feedback to them about the process, it can be experienced as extractive, disrespectful and disempowering.

Box 7: Risks of 'extractive' evaluation processes

Evaluations take people's time, and can often divert them from more productive activities, and they often are not the only data gathering processes communities participate in, which may include needs assessments and other evaluative processes conducted by other agencies as well. These processes often, however, do not report back to the communities in which they are implemented, or help communities understand how the information was used. As a farmer from a conflict-affected area of Sri Lanka observed:

> 'Before 1990, we were helping each other and the few organisations in our areas were listening to us. Now, it is different—all these foreigners and their assistant Sri Lankans who come in Land Cruisers with questionnaires only want our information. Then they disappear and a new group comes. I think that if they can't do any good, they shouldn't come" (as quoted by Jayawickrama & Strecker).

Sources: Jayawickrama, J. & Strecker, J. (forthcoming 2014) 'The Ethics of Evaluating Research—Views from the Field'. In Bush, K. & Duggan, C. Eds. *Evaluation in the Extreme: Research Impact and Politics in Violently Divided Societies.* New Delhi: Oxford University Press; Anderson, M., Brown, D. & Jean, I. (2012) *Time to Listen: Hearing People on the Receiving End of International Aid*. Cambridge, MA: CDA.

---

4. *Access to data and validity of the evaluation results.* Conflict can affect access to stakeholders or territories, require sudden adaptations to the design and conduct of an evaluation, increase levels of distrust and thus hamper data collection, or make follow-up studies difficult due to displacement of people and groups. These all potentially affect the validity of the evaluation findings. Moreover, when significant war-affected populations or significant stakeholders are excluded from an evaluation because of conflict developments, evaluators and evaluation managers need to consider whether the findings are ethically defensible, even with clear caveats, and whether they can accurately represent the intervention's impact[41]. They will also need to consider whether the evaluation's inability to include those perspectives might feed into a sense of grievance by neglected groups.

5. *Politicisation of evaluation and exacerbation of tensions.* Peacebuilding interventions take place in complex political economies where conflict parties and development partners alike have agendas and interests. In this context, the findings of an impact evaluation may be politicised, misinterpreted or selectively used by the conflict parties to pursue their agendas or to weaken their opponents. Any part of the evaluation process—from the questions guiding the evaluation to surveys, focus groups, and interview questions used to gather data, to the composition of the evaluation team—can potentially be seen as provocative or offensive and can be used by the parties to further their own arguments or can surface divisive issues, disappointed expectations or unaddressed grievances that can be lead to increased tension. This is true also of reporting and dissemination of evaluation results. The evaluation of Norway's role in the Sri Lankan peace process,

---

[41] Duggan, C. & Bush, K. (2014) 'Ethical Tipping Points in Conflict Zones'.

for example, was characterised by some Sri Lankan media as 'part of a larger international effort to undermine the government's victory over the LTTE and whitewash Norway's errors'[42].

3.5.2 Processes for managing conflict sensitivity risks

Many evaluators who conduct impact evaluations in fragile and conflict-affected contexts take a number of steps to ensure the evaluation process itself is ethical and conflict-sensitive.

Two broad strategies can assist evaluation commissioners and managers manage conflict sensitivity risks of impact evaluations in peacebuilding contexts:

1. *Knowledge, skills and experience to identify and manage conflict sensitivity issues.* Above all, this means ensuring that the evaluation team—both international and local members—possess a skill set and knowledge that go beyond the technical and methodological competencies required for evaluation generally. These include:

   - Knowledge and experience in the country and region itself allowing the team to filter out non-credible information, understand biases and agendas of different groups and actors, and understand the multi-layered dimensions of the conflict context and their implications for the evaluation
   - Political sensitivities, diplomacy and conflict resolution skills
   - Capacity to understand and potential to access a wide range of different perspectives and stakeholder groups in the conflict
   - Background and expertise in peace and conflict research and analysis frameworks
   - Cultural competence and cultural humility
   - Knowledge of the structures, strategies and behaviour patterns of armed actors, and
   - Anthropological, historical and political sensibilities[43].

2. *Processes to identify and deal with conflict sensitivity concerns.* Where potential conflict sensitivity problems have been identified, processes to deal with them can be integrated into the evaluation process. Evaluation consultation or advisory groups—ideally with local advisers and experts—could be a useful source of advice and support on conflict sensitivity of the evaluation design, evaluation team selection, methods, adaptation to changing conditions on the ground, as well as interpretation of data and presentation of findings[44].

> **Box 8: An Example: Processes to identify and deal with negative impacts of evaluation on conflict**
> In an evaluation of a Muslim-American virtual dialogue, the evaluator included a 'test' question at the end of each survey, asking participants to answer the question and provide feedback on their reactions to it and its appropriateness. This helped evaluators to avoid asking questions considered to be offensive or that risked reinforcing or recalling negative stereotypes.
>
> Source: Interview with evaluator.

---

[42] Goodhand, J. et al. (2014) 'Battlefields of method', p. 78.
[43] Adapted from Bush, K. & Duggan, C. (2013) 'Evaluation in Conflict Zones: Methodological and Ethical Challenges'; Duggan, C. & Bush, K. (2014 forthcoming) 'Ethical Tipping Points in Conflict Zones'.
[44] *Id.*

4 Identifying causes and assessing attribution and contribution—Choosing an appropriate evaluation approach

4.1 What is the 'attribution problem'?

A key feature of impact evaluations is that they do not just gather evidence about impacts, but also about the intervention's role in producing them. For peacebuilding especially, it is rare that causal attribution means total attribution, but rather it refers to partial attribution or analysing the intervention's contribution to impacts.

A range of methodologies exist for examining causality—whether pure attribution, partial attribution, or contribution. Each has merits and challenges associated with them. In all attempts to address the question of what caused what effect, there are three components that are commonly considered as important:

1. An investigation of the *factual*, i.e. what kinds of impacts occurred for whom? How did actual results of the intervention compare to those expected from the theory of change? Are results consistent with the theory that the intervention caused or contributed to them?
2. Exploring and ruling out or posing alternative explanations (other than the intervention) for the results.
3. Comparing the *counterfactual*, which either estimates or seeks to explain what would have happened if the intervention had not occurred.

This third component has conventionally been considered necessary for analysing attribution, and there has been debate about the best ways of establishing the counterfactual—given that it is not possible to observe how the situation would have been without the intervention. However, there is much debate about whether it is even possible to construct a good counterfactual in many peacebuilding interventions—given the complexity of fragile and conflict-affected contexts and the nature of the interventions, including policy interventions or interventions focused on macro-level effects.

Three broad categories of evaluation approaches exist:

Variable-based approaches identify a series of indicators that can be assessed through largely quantitative methods, and make judgements about causation based on comparing indicators before and after intervention with the same indicators in a situation in which the intervention *was not* implemented (i.e. a counterfactual). For this to be valid, other factors must be held constant, so that the comparisons are matching like with like. These include experimental and quasi-experimental methodologies and statistical modelling.

Process or mechanism-based approaches identify important activities or events, changes or impacts that occur, and seek to find, and analyse the processes that connect them. They elucidate how a result occurred, attempting to identify the one or more causal mechanisms (including interaction with contextual factors and other actors in the field), which can plausibly be argued to be relevant to the impacts. These approaches include theory-based evaluation designs and case studies.

Participatory approaches involve participants and stakeholders in the evaluation process as co-evaluators, not just sources of evidence. Programme participants are involved in determination of

objectives, indicators of success, methods to be used, as well as in data collection and analysis. Examples include Participatory Impact Assessment (PIA), Most Significant Change and Outcome Harvesting methodologies.

For each of these approaches, there tends to be debate about a) their ability to attribute impacts to any given intervention, and b) whether seeking sole attribution is a meaningful exercise in peacebuilding. Many of the challenges to attributing impacts to an intervention are described in Section 2.3, in particular trying to isolate the intervention from the many other factors contributing to a social change in any given environment. These challenges have led many practitioners and evaluators to focus on finding robust evidence of the intervention's *contribution* to the result, as causal links and assumptions in peacebuilding outcomes are seldom clear or direct. As a recent study commissioned by DfID noted, '…given the way development aid is understood nowadays we are assuming it more likely that "causes" will contribute to "impacts" rather than be more exclusively connected to them as the term "attribution" implies'[45].

4.2   The importance of mixing methods

As no design is without limitation, consideration should also be given to using more than one design to compensate for weaknesses in another and combining designs and methods in order to strengthen causal claims[46]. With broad consensus that causal inference is supported by integration of both quantitative and qualitative data, there is increasing use of mixed methods and combined or 'nested' designs for assessing causal attribution or contribution[47]. Triangulation is a key reason to mix methods, as it is about looking at things from multiple points of view. Methods can be combined in a number of ways, for example:

- Triangulating data—studying the intervention and the impacts using different types of data (e.g., quantitative and qualitative), different points in time or units of analysis.
- Triangulating researchers/evaluators and disciplines—having multiple evaluators, from different disciplines and different types of expertise (including country/local knowledge) look at the same problems.
- Triangulating methods—using different methods (e.g., surveys, focus groups, key informant interviews, observation, government statistics, etc.) and approaches (e.g., RCTs combined with ethnographic case studies, theory-based and participatory approaches) to fill in gaps and weaknesses each method has and provide compensate for weaknesses of different methods.

Data collected or results reached by one method can *corroborate* results reached by another or can *complement* the use of another to help understand and interpret results reached by one method. Thus mixing quantitative and qualitative methods is increasingly becoming a standard for robust impact evaluation[48]. Use of one evaluation approach (e.g., RCTs) in combination with another can also be useful in corroborating or providing further insights into evaluation findings. In both cases, the inferences and insights concerning impacts will be stronger and richer when combinations of methods are used.

---

[45] Stern, E. (2012) *Broadening the range of methods and designs for impact evaluation*, para. 2:30.

[46] *Id.*, p. 18.

[47] *Id.*

[48] See Bamberger, M. (2012) *Introduction to Mixed Methods in Impact Evaluation*. Impact Evaluation Notes No. 3. Washington, D.C.: Interaction. http://www.interaction.org/document/guidance-note-3-introduction-mixed-methods-impact-evaluation.

> **Box 9: Examples of mixing methods to corroborate and enhance findings**
>
> An impact evaluation of a reconciliation radio soap-opera, Musekeweya, in **Rwanda** used a group-*randomised* design, in which communities were randomly assigned to listen either to Musekeweya, or a health soap opera, both of which had outcomes relating to changing beliefs, norms and behaviours. The communities were sampled from categories representing relevant political, regional, and ethnic breakdowns of present-day Rwanda. Each community was matched to the most similar community according to a number of observable characteristics, such as gender ratio, quality of dwellings, and education level. Then, one community in each pair was randomly assigned to the reconciliation program and the other to the health program. In each community, a group of randomly selected adults (balancing for sex and age) listened to four 20-minute episodes once a month on a portable stereo with a researcher present. The researcher wrote an observation sheet after leaving, noting individual and group emotional reactions to content (observable enthusiasm, attentiveness, confusion, emotional expressions, and amount and type of discussion during and after the program. They recorded how often participants discussed program themes like intergroup prejudice, violence, or trauma. At the end of one year a team of 15 Rwandan researchers accompanied the 4 regular research assistants and team leader Paluck to each community for 3 days. They conducted individual interviews, focus groups, and behavioural observation with all 40 participants. Researchers also recorded group deliberations about how to share and supply batteries for the portable stereo and set of 14 cassette tapes of the radio program presented to each community at the end of the data collection. The qualitative data strengthened the evaluators' quantitative findings regarding the radio show's impacts, while also providing insight into how and why it had led to changes in social norms and behaviours but not personal beliefs.
>
> An evaluation of the cumulative effects of development assistance on stabilisation in northeastern ***Afghanistan*** used a ***quasi-experimental design*** to assess perceptions of communities on a number of issues (including security, the government, armed groups, etc.) combined with extensive background research in communities and case studies in atypical communities. The qualitative research provided important evidence for the conflict analysis, for understanding 'the quality and robustness of the causal model that is tested statistically' and for sound interpretation of the quantitative results.
>
> Sources: Paluck, E. (2009) 'Reducing Intergroup Prejudice and Conflict Using the Media: A Field Experiment in Rwanda'. *Journal of Personality and Social Psychology*. Vol. 96, No. 3, pp. 574-587; Bohnke, J., Koehler, J. & Zurcher, C. (2014) 'Assessing development cooperation in northeast Afghanistan with repeated mixed-method surveys'. In Andersen, O. et al. *Evaluation Methodologies for Aid in Conflict*. Chapter 6.

In choosing an approach or combination of approaches and methods, commissioners of evaluations need to weigh the advantages and limitations and the degree to which the limitations can be overcome or compensated for. The sections that follow (Sections 5-7) outline a variety of evaluation designs that can be used for evaluating impacts of peacebuilding interventions and consider their strengths and limitations, with a focus on three broad questions:

- Can the approach provide good answers to the evaluation questions of interest? Because different approaches are suited better to answer different kinds of evaluation questions, understanding what they can and cannot answer is important.
- What are methodological strengths of each approach, and what are the challenges and limitations? Can the challenges that threaten the validity of the evaluation results be addressed or minimised?
- What are conflict sensitivity risks, and can they be minimised or managed? Conflict sensitivity concerns relate both to the effects of the conflict on the evaluation and the effects of the evaluation on the conflict context.

5 Experimental and quasi-experimental designs: variable-based approaches

5.1 What are they?

Experiments and quasi-experiments are one of the main variable-based approaches. They seek to measure impact by measuring the difference in outcomes for those receiving the intervention (the 'treatment') compared with what the outcomes would be for the same people without the intervention. This is done by establishing a counterfactual or valid comparison group that resembles the participant group and context as much as possible. In *experimental methods* a large enough sample of people or units that share the same characteristics are divided in a way that is random, i.e. not self-selecting or selected through some other kind of bias or criteria. One group has access to the 'treatment'—i.e. participates in the programme or benefits from the policy— while those not selected (the 'control group') are not given access or are given access to a different programme.

In *randomised control trials (RCTs)*, the intervention must be designed in such a way to permit random assignment to a 'treatment' or 'control' group, through a method, such as a lottery, that prevents self-selection. This is intended to limit as far as possible the influence of other factors, or variables, that could affect outcomes. If the numbers of people or groups are large enough, statistically the 'treatment' and 'control' groups should be comparable. *Quasi-experimental* methods are used when it is not possible to randomly assign treatment. Other methods such as matching or delayed 'treatment' are used to establish groups that are as similar as possible to the treatment groups.

Evaluators assess and quantify what the impact of the intervention is by measuring the difference in outcome between the 'treatment' and 'control' or comparison groups before and after the intervention. The 'impact' is the net effect, expressed in % or numerical terms, averaged out across the group.

Box 10: Examples of experimental and quasi-experimental methods

Experimental and quasi-experimental designs have been used to evaluate the impacts of community-driven reconstruction programming in a number of fragile and conflict-affected states—from Sierra Leone and Liberia to Sudan, Congo and Aceh. They have also been used increasingly to study the effects of programming designed to affect attitudes (e.g., media, dialogue).

In an impact evaluation of *Tuungane, a* community-driven reconstruction programme in the Democratic Republic of Congo the evaluation team used *randomisation* to establish 'treatment' and 'control' groups for the intervention; the communities that received *Tuungane* were selected through public lotteries from a larger pool of potential participating communities. The evaluation sought to measure whether the programme 'improve[d]understanding and practice of democratic governance, citizens' relationships with local government, and social cohesion, and thereby communities' ability to resolve conflict peacefully'. It used household surveys alongside a behavioural intervention - an unconditional cash transfer grant to given to communities selected randomly to allocate to projects—to assess the difference in behaviour between Tuungane communities and non-Tuungane communities.

A *quasi-experimental* approach was used in an evaluation of the *Community-Based Reintegration Assistance for Conflict Victims* program funded by the Aceh Peace Reintegration Agency (BRA). The program sought to deliver quick assistance to conflict-affected villagers to improve their short-term material well-being, and to promote social cohesion, strengthen village decision-making mechanisms and enhance faith in government institutions. The evaluation used data from a household and village head survey conducted after the programme Because good records were kept on the criteria and process of selection of beneficiaries for the programme, it was possible to identify which non-beneficiary communities were similar enough treat them as a control group

Sources: Humphreys, M. Sanchez de la Sierra, R. & van der Windt, P. (2012) *Social and Economic Impacts of Tuungane: Final Report on the Effects of a Community Driven Reconstruction Programme in Eastern Democratic Republic of Congo.* New York: Columbia University and IRC; Barron, P., Humphreys, M., Paler, L. & Weinstein, J. (2009) *Community-Based Reintegration in Aceh: Assessing the Impacts of BRA-KDP.* Indonesian Social Development Paper No. 12. http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2010/03/31/000334955_20100331005320/Rendered/PDF/537140NWP0ISDP10Box345623B01PUBLIC1.pdf.

5.2   What kinds of peacebuilding impact evaluation questions is this approach useful for?

Because they are based on an established counterfactual, experimental and quasi-experimental approaches are considered by those who use them to be the *only* reliable means to show **attribution** of impact to an intervention, and to demonstrate clearly *whether* an intervention has caused a particular impact and, if so, by how much. This is largely a quantitative exercise. While this claim is challenged by others[49], experimental and quasi-experimental approaches have been used for:

- *Testing links in the chain of results in a theory of change.*   When the conditions for experimental and quasi-experimental designs (e.g. related to establishment and 'control' of the control or comparison group, statistical validity and significance of findings, etc.) can be met, they can provide evidence on whether one result in a hierarchy of results actually led to another, or whether a set of activities actually produced an expected specific change, ideally in conjunction with other methods (such as observation, interviews and focus groups). The impact evaluation of the Musekeweya radio soap opera (Box 9, above), for example, used an experimental design (combined with ethnographic and observational methods) to evaluate media influence strategies and impacts in promoting changes in personal beliefs about the programme's messages on prejudice, violence and trauma; in influencing social norms and in influencing behaviour. The evaluation provided 'some of the first clear evidence of the media's impact on intergroup prejudice and conflict'; it challenged commonly held theories that personal beliefs would be changed (and lead to behaviour change), finding instead that the programme influenced perceptions of social norms and behaviours based on these norms[50].

- *Providing evidence on what options might work better under particular circumstances.* They can test and compare results of alternative 'packages', alternative amounts of 'treatment', or alternative kinds of interventions for achieving the same goal in a specific context—for example different packages for demobilised combatants for whom all other variables are the same, or different lengths and intensity of training. In the World Bank's impact evaluation of the National Solidarity Programme (NSP) in Afghanistan, the evaluation not only compared treatment groups (those who received NSP with those that did not), but also introduced a randomised variation in the method by which projects were selected by 'treatment' communities to test whether the procedures made a difference to results within those villages. Half of the 'treatment' villages selected projects by referendum and half by community meeting. Evaluators also tested different methods (through randomisation) by which the community development councils were elected (by neighbourhood or 'at large')[51].

---

[49] *See* Stern, E. et al. (2012) *Broadening the Range of Designs and Methods for Impact Evaluation*; Piccioto, R. (2012) 'Experimentalism and Development Evaluation: Will the Bubble Burst?' *Evaluation*. Vol. 18, No. 2, pp. 213-229.

[50] Paluck, R. (2009) 'Reducing Intergroup Prejudice and Conflict Using the Media: A Field Experiment in Rwanda'. *Journal of Personality and Social Psychology*. Vol. 96, No. 3, p. 578.

[51] Beath, A., Christia, F. & Enikolopov, R. (2013) *Randomized Impact Evaluation of Afghanistan's National Solidarity Programme*. Washington, D.C.: World Bank. p. 8.

- *Validating proof of concept.* RCTs have been used in development and medical sciences as one methodology that helps validate proof-of-concept studies. When there is a growing body of evidence that suggests that a theory of change is or may be working, an RCT can be commissioned to validate parts of the theory of change. If there is growing evidence, through case studies, evaluations, and ethnographic studies, and the RCT produce similar results; then the RCT is able to confirm or validate the level of effect in that particular situation. The experimental methods may also be able to test sequence of activities and see which sequence works best to bring about sustainable change. Multiple RCTs in different locations and contexts, along with combination of on-going evidence gathering, may be able to point out which interventions are proven solutions and should be brought to scale given the conflict dynamics.

5.3   What kinds of peacebuilding impact evaluation questions are these approaches not useful for?

Experimental and quasi-experimental methods have limitations that make them inappropriate for many, if not most of the peacebuilding impact questions that may be of interest to practitioners and policy makers. They do not provide answers on key questions about peacebuilding impacts, including:

- *Effects on 'peace writ large' or drivers of conflict.* As experimental and quasi-experimental methods work better in short results chains (direct results, without many intervening steps or changes), where conditions can be controlled, they are not good for assessing the effects of peacebuilding interventions on drivers of conflict. They cannot assess the collective changes in societal attitudes nor whether those changes have affected tensions, violence or other dimensions of conflict or peace.

- *Effects on policies, culture change, behaviour of a single institution, or political and legislative processes and outcomes, mediation and 'track 2' dialogue.* For many of these peacebuilding interventions, there are no meaningful control or comparison groups. If an objective of a programme includes policy change of on a critical driver of conflict such as land reform on one institution, it is of no use to compare that to another institution that doesn't get the 'treatment' This holds for mediation work, advocacy work (especially at the national level) assessing the influence of work with specific group(s) on inter-group perceptions and stereotypes on collective norms and behaviours and many institutional reform interventions (justice reform, civil service reform, etc.).

- *Unintended and unanticipated consequences.* Experimental and quasi-experimental approaches can only establish credible comparisons for planned results. As unintended or unexpected results are unforeseen, they cannot be incorporated into these processes[52]. In particular they have no utility for evaluating unintended impacts (positive and negative) of interventions on conflict—an important question both for peacebuilding and development interventions in fragile and conflict-affected contexts.

---

[52] In the impact evaluation of the *Tuungane* community-driven reconstruction programme in Congo (see Box 10 above), the evaluation team asked whether the project generated jealousies with other villages; one-third of respondents answered that it did. While the evaluation did not gather any data on effects on *existing* dividers or connectors, it was able to detect some unintended impacts. The evaluation did not explore them further; went on to note that '[a]lthough we flag this concern over discord, we see here that overall these patterns suggest that the project was well received and that the approach to development was appreciated'. Humphreys, M. et al. (2012) *Social and Economic Impacts of Tuungane*, p. 20.

- *Assessing how and why an intervention may or may not have made a difference.* Experimental designs in isolation are not intended to determine the reasons why and how an intervention may have produced certain effects. They are largely they are concerned with quantifying whether they did or did not have impact and how much of an impact, or effect, they caused. Unless they are combined with other approaches they cannot address key evaluation purposes for peacebuilding such as:

  - How to go about improving or adapting an intervention – this requires evidence to understand what changes might be needed, where and why.
  - Whether and how an intervention might be replicated elsewhere—this requires an understanding of what contextual or other factors are important for the success of an intervention.
  - Whether it is possible and how to scale up a pilot initiative—this requires an understanding of the relationship between context, through conflict and relevance analysis, and the programme.

  Relying *only* on results of impact evaluations that use an experimental or quasi-experimental approach to make decisions regarding adaptation, continuation, replication or scale-up of a peacebuilding intervention, could lead to policy mistakes that undermine effectiveness.

- *How an intervention impacted different groups differently.* Because experimental and quasi-experimental approaches are methodologically designed to average out the *net* effect of an intervention across different groups, they are weak in identifying differences that might relate to important variables. In conflict situations, where horizontal inequalities can drive tensions, understanding the range of effects—for whom the intervention had and did not have impacts—it is important not to average out results, as important conflict-related evidence would be masked.

> **Box 11: The need to understand differentiated impacts**
> The evaluation of IRC and CARE's *Tuungane* community-driven reconstruction programming in DRC, (described in Box 10 above) found 'no overall impact' of the programme. At the same time it noted, 'It is possible that there are positive effects for some and negative effects for others, with small or no effects for the average'. This lack of nuanced data on who experiences positive or more negative outcomes (and under what conditions) is needed to allow implementing teams or DFID to improve programming. One comparative study of impact evaluations of community-driven development and reconstruction programming done using experimental designs notes that programme teams and participants find it hard to recognise their work reflected in evaluations when results are averaged out to 'zero'.
>
> Sources: Humphreys, M. Sanchez de la Sierra, R. & van der Windt, P. (2012) *Social and Economic Impacts of Tuungane: Final Report on the Effects of a Community Driven Reconstruction Programme in Eastern Democratic Republic of Congo*. New York: Columbia University and IRC, p. 75; King, E. (2013) *A Critical Review of Community-Driven Development Programmes in Conflict-Affected Contexts*. London: DfID and IRC. http://r4d.dfid.gov.uk/pdf/outputs/misc_gov/61046-A_Critical_Review_of_CDD_in_Conflict_Affected_Contexts.pdf.

5.4 What are methodological strengths of this approach in evaluating peacebuilding impacts?

There are a number of requirements that need to be met for a robust and valid experimental or quasi-experimental design, such as identification and 'control' of control or comparison groups, prevention of 'contamination' or 'spillover' effects between treatment and control groups, sufficient numbers for statistical significance, etc. These are strict and (as discussed below) can be difficult to meet. However, if they are met, a strength of experimental and quasi-experimental methods relates to their ability to:

- *Provide data tailored to the evaluation question.* Methodologically, because experimental and quasi-experimental designs are undertaken from the design phase of an intervention, they ensure that a baseline is conducted and that data that is relevant to the questions being evaluated collected and available. This is in contrast to evaluations begun after-the-fact, where the evaluators may need to rely on a programme's monitoring data and baselines (which may be weak, or non-existent), or need to reconstruct baselines.

- *Address issues of bias that can distort evaluation findings.* Experimental and quasi-experimental methods are also considered by many to be rigorous in reducing *'selection bias'*, thus ensuring accurate evaluation findings. Selection bias refers to the potential to skew results of the evaluation when participants in a program are systematically different from non-participants (even before they enter the program). Bias can affect the legitimacy of impact evaluations, and in particular, where most impact evaluations seek to compare program participants to non-participants in order to infer the effect of the program, this selection bias can skew evaluations in a positive or negative direction. While all methods are subject to this problem, experimental and quasi-experimental methods, are particularly good at reducing the risks of this kind of bias, as control groups established through random assignment or statistical method permit as close an approximation as is possible of what would have happened without the intervention.

5.5   What are methodological challenges and threats to data validity of this approach for evaluation of peacebuilding impacts?

There are a number of serious practical challenges that make experimental and quasi-experimental designs inappropriate for many peacebuilding interventions and that undermine the credibility and validity of the evaluation results:

- *Dealing with complexity.* Experiments and quasi-experiments work largely where stability and controllable conditions enable examination of one primary cause and one (or few) primary impacts. In complex interventions in unstable situations in which the context and interactions of multiple causes are important for understanding impacts, or where there is overlap with other interventions and actions with similar aims, the variables are too many to be 'controlled'.

Box 12: A framework for identifying 'complexity'
Michael Woolcock has developed a framework for identifying the 'causal density' or complexity of development interventions for purposes of developing evaluation strategies and choosing methods. These are relevant to peacebuilding as well. Elements include:

- Is it transaction-intensive (i.e., many person-to-person interactions)?
- Is discretion or local adaptation required for implementers on the ground?
- Is there pressure on implementers to do something other than respond productively to the situation on the ground?
- Do implementers have to innovate in situ vs. deploy options from a known menu?

The more these elements are present in operational aspects of the programme, the more 'complex' the intervention, and the more useful mechanism-based methods such as case studies are.

Source: Woolcock, M. (2013) 'Using case studies to explore the external validity of 'complex' development interventions'. WIDER Working Paper No. 2013/096. UNU-WIDER.

- *Identifying control or comparison groups and 'controlling' them.* Attitudes, behaviours, perceptions and other 'invisible' dimensions that are central to peacebuilding can be easily masked, rendering apparently equivalent control or comparison groups different from treatment groups in salient ways. Isolating the 'control' or 'comparison' group completely from the 'treatment' group, to prevent 'spillover' effects, is not only very difficult in conflict contexts, it also may run counter to the objectives of the intervention itself. Where other programme interventions are working with the same groups of people, or where synergy of approach with other key actors is important, creating control and treatment groups for one intervention is counter-productive.

Box 13: Randomisation may not be 'random'

In Liberia the randomisation of a community-driven reconstruction programme resulted in treatment areas being more likely to be rural and located farther from health clinics than control area. If urban/rural divides are relevant to the conflict, or different ethnic groups live in different areas, this can obscure or compromise findings in relation to peacebuilding.

Source: Fearon, J., Humphreys, M. & Weinstein, J. (2008) *Community-Driven Reconstruction in Lofa County: Impact Assessment.* New York: IRC. http://www.columbia.edu/~mh2245/FHW/FHW_final.pdf.

- *Keeping the intervention the same.* Experimental and quasi-experimental designs require that the intervention essentially be kept constant throughout the time period of the evaluation (and the programme), so as to maintain a valid comparison. However, in peacebuilding contexts, interventions should be led by the demands of the context. Conflict is dynamic and the conflict drivers may change during the life of a programme. There is a risk that either the carefully designed experiment would become less relevant, or only able to measure pre-selected indicator that have been superseded by events, or by the time of the data collection the indicators and mode of implementation may well have been adjusted to meet new circumstances.

- *Maintaining sufficient numbers and balance.* Drop-out rates from treatment or control groups (for example due to conflict-induced displacement) can undermine the statistical significance of experimental or quasi experimental approaches. The statistical power of the evaluation findings depends in part on the numbers; thus attrition rates –treatment or control participants that leave or cease participating for a variety of reasons—are important. These are usually expected in any experimental or quasi-experimental evaluation, but can be higher in conflict areas, where conflict might lead to displacement, intimidation resulting in changes of attitude toward participation in the programme or evaluation, etc. In the evaluation of the *Tuungane* community-driven development programme in Congo (Box 10 above), data was collected successfully in 78% of the villages and 62% of individuals, with significant loss of data because of the expulsion of one of the evaluation teams from a province due to political tensions and the inaccessibility of other areas due to security developments[53]. In Afghanistan (NSP Programme, Box 4 above) attrition rates for household surveys during the evaluation of the National Solidarity Programme ranged from 9% to 28%. This raises the possibility that an evaluation may not be able to generate valid findings, or, even when attrition does not affect findings in a statistically significant way, whether voices being left out of the evaluation are significant enough to compromise the integrity of the findings.

---

[53] Humphreys, M. et al. (2012) *Social and Economic Impacts of Tuungane*, p. 39. In both Congo and Afghanistan, the loss of data resulting from attrition and inaccessibility was not statistically significant, in part due to the large numbers and how attrition was spread between treatment and control areas. However, the risk remains high.

- *Ensuring that indicators of peacebuilding outcomes accurately measure the underlying realities.* This is the question of 'construct validity', and it is a challenge in all impact evaluations. However, in peacebuilding it is more difficult. Peacebuilding outcomes (e.g. governance, reconciliation, trust, cohesion) are often described as 'hard-to-measure' in quantitative terms, which is where experimental approaches focus their attention. A challenge relates to whether the indicators used can adequately 'measure' or represent such outcomes. In order to quantify the amount of 'social capital' generated by a particular intervention, for example, some evaluations have used measures such as people's willingness to lend money to another community member, or a migrant, or asking people to decide whether to lend nominal amounts of money to public projects, or decide how to use a small gift. These may not provide insight into dimensions of social capital that are relevant to conflict and peace—inter-group trust, willingness to live together, or level of inter-group grievances[54].

  Moreover, careful attention needs to be paid as to whether measures capture non-linear and multi-dimensional dynamics of change. If for example a programme to reduce gender-based violence uses reports of violence as a measure, it may not capture what is actually happening. Reporting of gender-based violence is known to be affected by numerous cultural and other factors; an increase in reports may not reflect an increase in actual gender-based violence. This may in fact have decreased in 'treatment areas' relative to control, but the measure cannot capture that dynamic[55].

- *Timeliness.* Experimental approaches often require the involvement of large data collection exercises at fixed points in time, and the expertise of highly-qualified statisticians to analyse and compile data-sets in ways that can take significant amounts of time. The complexity of variables inherent in conflict-affected environments, and the way interventions rightly seek to affect several interlinked objectives, means that the RCT approach alone is unlikely to be able to answer the complex questions that need to be addressed in any meaningful evaluative exercise. Indeed, the evaluation of the *Tuungane* community-driven reconstruction programme in Congo asked this question in its 'speculation' regarding its finding of null effects; the timing may have been too soon to have seen economic or other effects[56].

5.6   What are special conflict sensitivity concerns associated with this approach? Can these be addressed?

Evaluators using RCTs have found ways to address ethical concerns regarding denial of access to benefits of the programme due to the evaluation process (randomisation). In some cases, they take advantage of the fact that a programme may be staged, and identify the 'control' groups among people and communities who are not receiving the programme *now* but will later. This was the case in an evaluation of ex-combatant reintegration in Burundi, where it was not possible to deny benefits to some ex-combatants altogether[57]. Others have used randomisation as a fair process for distributing limited resources, when it is clear that not all potential participants in a programme will be able to participate.

---

[54] See *id.*, p. 62. The evaluation used people's willingness to lend money to other village members as measure of social cohesion.

[55] Glennerster, R. & Takavarasha, K. (2013) *Running Randomized Evaluations: A Practical Guide*. Princeton: Princeton University Press, p. 89.

[56] Humphreys, M. et al. (2012) *Social and Economic Impacts of Tuungane*, p. 76.

[57] Samii, C., Brown, A. & Kulma, M. (2012) 'Evaluating Stabilization Interventions: Working Draft 2.0'. 3ie White Paper. https://files.nyu.edu/cds2083/public/docs/evaluating_stabilization_interventions_120816shortenedb.pdf;

Risks of conflict insensitivity include but go beyond these ethical considerations and should be examined carefully by evaluation commissioners and managers. In addition to the general concerns about access to data and impacts of data collection methods on conflict, bias, and political consequences of evaluation processes, described in Section 3.5 above, a number of elements of experimental and quasi-experimental methodologies have the potential inadvertently to interact negatively with the conflict:

- *The use of control groups.* Transparent, 'lottery' style processes for allocating benefits and simultaneously randomising 'treatment' and 'control' groups have at times been considered fair and prevented potential conflict. This was the case in the CARE-IRC *Tuungane* community-driven reconstruction programme in DRC, where control and treatment groups were allocated through a transparent lottery process—and 75% of people surveyed in non-Tuungane areas during the evaluation believed the process was fair[58]. In other circumstances, however, where causes of conflict or tensions involve lack of access to resources or discrimination, making a control group aware of its relative deprivation (lack of resource inputs) may result in violent mobilisation, particularly if the lines of division in the conflict overlap with the selection criteria for targeting.

- *Behavioural games.* Behavioural games have become increasingly popular methods to measure levels of trust and social capital in experimental and quasi-experimental methods. However, they risk escalating tensions and undermining capacities for peace within communities. In general, behavioural games are a poor substitute for qualitative or ethnographic approaches, as they can certainly state that certain things occurred, but they are lack the vital information to interpret the results in contextually and culturally relevant ways.

Box 14: Potential risks of behavioural games

The impact evaluation of the *Tuungane programme in DRC* (described in box 10 above) used a behavioural game to gain data on a number of measures related to governance, transparency, social cohesion, among others. It selected 560 villages—half chosen randomly from the 'treatment' and the other from the 'control'—to participate in an unconditional cash transfer programme in which they were told they would receive grants of $900 to be used on projects to benefit the village. (In fact, $1000 was given to provide a measure of whether leaders would report unanticipated gains to the villagers.) The evaluators then observed the interactions in relation to this grant: who came to the meetings, who was chosen to manage the funds, whether accountability procedures were put in place, transparency of the process, capture/who benefitted most, etc. Given that poor governance is a cause of conflict in this context, this behavioural game risked actively aggravating a known cause of tension.

In **Sudan**, a 'standard dictator game' was implemented as part of an impact evaluation of a community-driven development programme; community members—'laboratory subjects', as they were called—were given three Sudanese pounds and asked to decide how much, if any, they would donate to a local needy family. Notably by treating people as human subjects rather than agents, this 'game' risked reinforcing perceptions of disempowerment and disrespect of community members, and providing very visible and widely observed acts of marginalisation and deprivation of resources (which remain without remedy).

Sources: Humphreys, M. et al. (2012) *Social and Economic Impacts of Tuungane*, p. 29; Avdeenko, A. & Gilligan, M. (2014) *International Interventions to Build Social Capital: Evidence from a Field Experiment in Sudan*. Policy Research Working Paper 6772. Washington, D.C.: World Bank and Development Research Group.

Gilligan, M. Muukiyehe, E. & Samii, C. (2013) 'Reintegrating Rebels into Civilian Life: Quasi-Experimental Evidence from Burundi'. *Journal of Conflict Resolution*. Vol. 57, No. 4, pp. 598-626.

[58] Humphreys, M. et al. (2012) *Social and Economic Impacts of Tuungane*, p. 26.

6 Theory-based and case-based designs: Mechanism-based approaches

6.1 What are they?

Theory-based approaches focus on the underlying assumptions in an intervention's design—how activities will lead to outcomes, and how outcomes will have broader contributions to peace—and make them explicit and test them. These approaches seek to identify why and how changes have taken place, discovering the causes of observed effects by identifying and analysing the *mechanisms* by which those effects happen. Case-based approaches focus on rich description of a single phenomenon or event –for example a programme, issue, institutional development, or event. Theory-based and case-based approaches use both qualitative and quantitative methods to identify trends, triangulate analysis, and analyse effects and mechanisms.

Theory-based methods have been used extensively in impact evaluations of peacebuilding interventions, from large multi-donor evaluations to programme and project evaluations seeking to assess impacts on higher-level objectives such as policy reform, and on drivers of conflict. In peacebuilding, theories of change can be difficult to identify and reconstruct, not only because of staff turnover in agencies, but also because of the political nature of peacebuilding work as well as the geopolitical and other agendas of donors that make development of a consistent and coherent theory of change difficult. In those cases, evaluators have developed 'working hypotheses', rather than complex hierarchies of results, based on documentation and interviews with stakeholders concerning the implementation of the intervention(s).

There is a great variety of tools and designs that are including under the general heading of theory-based approaches. Following is an illustrative sample, not exhaustive, of theory-based designs.

6.1.1 Realist evaluation

'Realist evaluation', first developed by Pawson and Tilley[59], seeks to answer questions of what works for whom, and in which contexts or circumstances the intervention is implemented. It seeks to answer questions about what causal mechanisms are triggered by which elements of interventions in what contexts? It identifies key change mechanisms (theory of change), contextual factors that influence the intervention or how it affects people, and outcome patterns, including how they differ for different groups of people. The 'CMOC package' (causal mechanism-context-outcomes configuration) is examined and subjected to 'systematic tests' using data collected to see if the model explains the 'complex footprint or outcomes left by the programme'[60].

6.1.2 Contribution analysis

Contribution analysis explores the contribution an intervention is making to observed results by verifying the theory of change behind it and assessing the contribution of other factors. It develops a 'contribution story' through identification of and evidence gathering about the logic of the links in the theory of change. Causality is inferred from: the reasonableness and plausibility of the theory of change (e.g., are the assumptions on which it is based sounds, and is it agreed by a range of key actors?); demonstration that activities were implemented; verification that the chain of results anticipated by the

---

[59] Pawon, R. & Tilley, N. (1997) *Realistic Evaluation*. London: Sage.
[60] Pawson, R. & Tilley, N. (2004) 'Realist Evaluation'. Paper funded by British Cabinet Office. http://www.communitymatters.com.au/RE_chapter.pdf.

theory occurred or didn't; and identification of other influencing factors and assessment of their degree of influence.

6.1.3   Process tracing
Process tracing uses qualitative analysis focused on analysing trajectories of change and causation to uncover the causal mechanisms for observed results. Process tracing attempts to explore the linkages between plausible causes and outcomes by tracing and comparing the sequence of events constituting the process of change and subjecting them to evidence tests to eliminate rival explanations.

6.1.4   Case studies
Case studies focus on a particular unit—a site, group of people, intervention, institution or event - and describe the case 'as a whole and in its context'[61], including identifying antecedents, examining what happened, in what timeframe, and with what consequences. They focus intensively on the causal mechanisms that lead to the impacts of interest to provide a rich understanding of the intervention in context and give voice to beneficiaries. Single case studies examine the congruence of the case with theory and counterfactual reasoning to determine causation, while comparative case studies use the logic of controlled case comparison, where cases are intentionally selected to provide factual variation.

---

[61] United States General Accounting Office, Program Evaluation and Methodology Division (1990) *Case Study Evaluations*. GAO/PEMD-91-10.1.9. Washington, DC: GAO. http://www.gao.gov/special.pubs/10_1_9.pdf.

Comparative case studies can examine multiple or connected causes of the same outcome, and cumulative case studies bring together findings from many cases to answer evaluative questions[62].

Box 16: Examples: Uses of case studies for impact evaluation

In the joint evaluation of the impacts of *conflict prevention and peacebuilding assistance in Democratic Republic of Congo* (see Box 3 above) the evaluation team used in-depth case studies to identify critical outcomes rapidly, enable the team to study projects deeply in their socio-political context, and enable exploration of contested viewpoints and multiple perspectives that characterised the conflict. Methods to identify 'net change' (e.g. quasi-experimental methods) were ruled out because of lack of baselines and indicators, and because conditions had 'changed so dramatically to make before/after or with/without comparisons meaningless'. Case studies were chosen purposely to explore to the greatest degree possible critical themes that had been identified in policy, conflict and portfolio analysis as well as interviews.

A 2006 evaluation of the impacts of the *Kecamatan Development Programme (KDP)*, a community-driven development programme, on local conflict in Indonesia integrated case studies with other methods to draw its findings on impact. The evaluation team conducted nine months of qualitative fieldwork, including 68 case studies of conflict pathways (exploring the evolution of a wide range of specific conflicts), key informant surveys in 'treatment' and 'control' villages, ethnographic study of key themes, and analysis of 'large-n' quantitative surveys. Selection of sites for case studies was done to include diverse provinces (to test whether patterns were present across contexts), and within provinces to compare 'high capacity' (for conflict management) with 'low capacity' areas. Fieldwork was conducted in sub-districts covering three 'treatment' and one 'control' site chosen using statistical methods to match on observable variables. This allowed the team to assess impacts on local conflict and conflict management and develop rich explanations of how and why KDP had (little) effect on conflict at the aggregate level, yet did help improve inter-group and state-society relations.

Sources: Brusset, E. et al. (2011) *Amani Labda (Peace Maybe): Joint Evaluation of Conflict Prevention and Peacebuilding in the Democratic Republic of* Congo. Brussels: FPS, Foreign Trade and Development Cooperation; Barron, P., Diprose, R. & Woolcock, M. (2006) *Assessing the Impact of the Kecamatan Development Program*. Indonesian Social Development Paper No. 10. Jakarta: World Bank.

6.1.5 Methods for ruling out alternative explanations in theory-based, case-based and participatory methods—when there is no counterfactual

When no counterfactual is possible, a number of methods exist for identifying and ruling out alternative possible explanations in theory and case-based designs, an important element of determining whether and how an intervention contributed to impacts. Methods for doing so include:

- *General elimination method (GEM).* The GEM identifies possible alternative explanation for impacts of interest and systematically investigates them to see if they can be ruled out. It has two stages: 1) identifying possible explanations that the intervention produced the impacts through key informant interviews, previous evaluations and research, and brainstorming; 2) gathering data to see if the alternative explanations can be ruled out. The underlying belief is that each of these Possible Causes would have a set of identifiable events or occurrences (*'modus operandi'*) that would indicate that cause was valid. Any cause or theory where the modus operandi has not been observed can be eliminated, leaving one explanation with a true causal link.

- *Searching for disconfirming evidence and following up exceptions.*

---

[62] Khagram, S. et al. (2009) 'Evidence for development effectiveness'. *Journal of Development Effectiveness*. Vol. 1, No. 3, pp. 247-270.

- *Multiple lines and levels of evidence (MLLE) and collaborative outcomes reporting.* These approaches build in a process for review of the evidence collected by a panel of credible experts, who check consistency of the evidence with the theory of change, check the credibility of the evidence regarding what impacts have occurred and the extent to which the evidence supports attribution to the intervention, and identify and explain exceptions. They might, for example, look at the strength of the evidence, its specificity, whether the evidence of impacts is too soon or too late to be attributable to the intervention, coherence with other accepted evidence, plausibility. In collaborative outcomes reporting, in addition to expert review, community consultations are used to check the evidence.

6.2 What kind of peacebuilding impact evaluation questions is this approach useful for?

Theory-based and case study approaches offer a number of advantages for impact evaluations of peacebuilding interventions. While they do not address sole attribution of impacts, they are effective at developing robust contribution stories. As a recent study commissioned by DFID noted, '…given the way development aid is understood nowadays we are assuming it more likely that "causes" will contribute to "impacts" rather than be more exclusively connected as the term "attribution" implies'[63]. In this sense they are able to provide a more realistic assessment of causality in complex, multi-dimensional settings, as they acknowledge that other factors are at play and seek to understand them.

Theory-based and case study approaches answer many of the evaluation questions that are of interest for peacebuilding practitioners and policy makers, and are often the only way to answer questions of impact adequately, such as:

- *How and why an intervention caused impacts.* At the core of theory-based approaches is the observation and tracking of *real* causal mechanisms and processes. They are therefore particularly well-suited for evaluations whose purpose is learning and programme improvement or redesign. They help answer what went right or wrong and why, and how important were contextual and other factors were in the results, and generate insights critical to adaptation and revision of implementation designs, for example:

    o Understanding whether failure to have expected impacts is due to a flawed (or incomplete) theory of change, or to implementation challenges or failures or flawed (or incomplete) analysis of the conflict
    o Understanding what obstacles and facilitating factors to future success need to be addressed or accounted for, and
    o Understanding whether and under what conditions an intervention might be replicable.

- *Contribution to policy, institutional reform, culture change, political processes, mediation and other high level change.* Theory-based and case-based methods, along with participatory approaches, are the primary way to understand impacts on policy (e.g., advocacy interventions). These approaches are particularly relevant for evaluating impacts of peacebuilding interventions, which often involve questions of policy reform, institutional reform or transformation in cultural norms with impacts at a collective (e.g. national) level.

---

[63] Stern, E. et al. (2012) *Broadening the range of methods and designs for impact evaluation – report of a study commissioned by the Department for International Development*, para. 2.30.

- *Peace writ large and broader effects of interventions.* Theory and case-based approaches are particularly appropriate for evaluating impacts that lie beyond the immediate sphere of influence or participants of the intervention. This is because they examine causal pathways and are able to assess how intervention-level outcomes affect broader conflict dynamics, or how development interventions with development objectives such as income generation, judicial reform, or police reform have affected conflict dynamics.

- *Evaluating unintended and unanticipated impacts.* The broader focus of theory- and case-based methods, especially case studies, in examining how causal factors combine, make them better suited to understanding how interventions may have contributed (or not) to unintended outcomes. Theory-based methods may not help to identify what unintended impacts on conflict may have occurred, but they can help determine whether and how an intervention may have contributed to them. They are thus particularly relevant (along with participatory methods) for evaluating **conflict sensitivity**[64].

- *Evaluating sustained impacts.* Assessment of long-term, sustained impacts, and the intervention's links to them is complicated by difficulties of understanding how external influences (including contextual factors and other subsequent interventions) as well as feedback loops have influenced results.

6.3   What kinds of peacebuilding impact evaluation questions are these approaches not useful for?

Theory and case-based methods have ways to assess the significance of a contribution, including some numeric techniques that support interviewing people for weights or degrees of contribution, but generally are not able to generate precise estimates of attribution. When impact evaluations are being conducted for purposes of accountability, especially to donors, therefore, mechanism-based approaches may provide weaker answers than experimental or statistical designs when both are feasible to do.

6.4   What are the methodological strengths of these approaches in evaluations of peacebuilding impact?

One of the important strengths of theory and case-based approaches is that they deal well with complexity. In fragile and conflict-affected contexts, any intervention or group of interventions rarely is the only factor influencing a particular outcome, and often has impacts only in conjunction with other contextual factors and interventions. For example, where years of dialogue programming and programmes encouraging cross-conflict cooperation helped prepare the ground for peaceful elections, and international monitoring and support to an election commission was the more proximate 'cause' of peaceful elections, theory and case-based approaches can identify the role of the dialogue and cooperative project interventions in the package and identify how the various interventions worked together.

Theory-and case-based approaches have advantages with regard to two additional methodological issues affecting the validity of the evaluation results:

---

[64] Methods for monitoring and evaluation of conflict sensitivity are covered in detail in another Guidance product of CCVRI: Goldwyn, R. & Chigas, D. (2013) *Monitoring and Evaluating Conflict Sensitivity*.

- *'Construct validity'.* Construct validity refers to the degree to which the evaluation measures measure what they purport to be measuring. For example, in the evaluation of the NSP in Afghanistan, incidence of disputes and resolution rates, as well as level of interpersonal trust were used as measures of social cohesion[65]. The issue of 'construct validity' would entail asking whether these measures in fact reflect dimensions of social cohesion in a conflict context. For example do attitudes towards migrants and willingness to lend money to returnees, migrants or ex-combatants reflect social cohesion that is relevant to conflict[66]? Peacebuilding interventions often set ambitious, often difficult to measure goals –such as reducing violence and fear, or accountable and inclusive governance—theory and case-based approaches are often better able to provide accurate measures of what those concepts mean in the particular context than are variable-based approaches, as they employ multiple ways of understanding these phenomena and the ways interventions affect them, as well as deep understanding of context.

- *'External validity'.* External validity refers to the degree to which the findings of an evaluation can be generalised or extended to other settings. Can one infer, for example, from the results of an impact evaluation of ex-combatant reintegration in Burundi that similar results will be obtained if the programme is implemented elsewhere, or if it is scaled up or replicated in Burundi? Experimental and quasi-designs are weak on this point; while they might be able to determine that the reintegration programme increased satisfaction with civilian vs. combatant life, they cannot say with certainty what will happen if the same programme is introduced into a different context. Theory and case-based approaches are not inherently strong for drawing conclusions about whether an intervention works elsewhere; however, their capacity (especially case studies) to untangle contextual and other factors, as well as implementations factors, that influence outcomes gives them a stronger potential to do so.

6.5   What are methodological challenges and threats to data validity of these approaches for evaluation of peacebuilding impacts?

The most significant criticism of theory and case-based approaches is that they cannot deal adequately with **bias** that may result in over-or underestimation of the intervention's role in contributing to a result. This includes bias resulting from the fact that programme participants may differ from non-participants because of characteristics that the evaluator cannot observe or from bias in sampling of whom to talk to or what cases to study (*'selection bias'*, see Section 5.4). For example, the participants may be 'easy to reach', i.e. have more favourable attitudes towards peacebuilding than non-participants, and thus be more likely to be impacted by the intervention. Or they may have different experiences of the conflict or its consequences that lead them to have different motivations and different stakes in the intervention.

Where **narrative analysis** (in contrast to categorisation and comparison of factors) is used as an analysis strategy there is a risk that they may not be specific enough about causality, confuse chronological sequences with causation, and avoid rigorous examination of evidence and conclusion because of the

[65] Beath, A., Christia, F. & Enikolopov, R. (2013) *Randomized Impact Evaluation of Afghanistan's National Solidarity Programme—Final Report*. Washington, DC: World Bank and MIT. p. 80.

[66] These were measures used in the impact evaluations of community-driven reconstruction programming in Democratic Republic of Congo and Liberia, among others. *See* Humphreys, M. et al. *Social and Economic Impacts of Tuungane*; Fearon, J., Humphreys, M. & Weinstein, J. (2008) *Community-Driven Reconstruction in Lofa County: Impact Assessment*. http://www.columbia.edu/~mh2245/FHW/FHW_final.pdf.

compelling 'storytelling' dimension of the methods[67]. However, While these are manageable, and narrative analysis can be very useful in elucidating causal processes, especially if they are complex and dynamic, evaluation commissioners and managers should be aware of the issues and seek to address them in drawing conclusions.

While bias is a challenge in all evaluation designs, the critique is that it is potentially worse in impact evaluations using these methods. As Howard White, Executive Director of 3ie notes:

> The possibility of demonstrating causality through other means is not open to question. The points where critics, in my view, need to give some ground is to accept that: (1) there is a problem of selection bias which needs addressing in any assessment of attribution; (2) RCTs are very often the most appropriate means of dealing with this problem; and (3) if not RCTs, then some other quantitative method must be used[68].

While there has been debate about whether RCTs in fact produces equivalent groups, approaches contribution analysis, or case studies need to take extra measures to minimise these risks of bias, as the approaches are not inherently bias-prone. The critique stems in part from the fact that theory-based and case-based approaches remain the primary approaches used in impact evaluations (including of peacebuilding) and are not often designed or implemented rigorously—either because of budgetary constraints or because of gaps in methodological expertise. Implementing quality theory-based and case-base methodologies in conflict and fragile environments is difficult and challenging. It is important when hiring evaluators and research teams that the individuals commissioned to undertake these evaluations have been well trained and have the appropriate academic and/or professional qualifications. In particular, evaluators should be able to speak to:

- *Sampling methodology* is just as important in qualitative research as it is in quantitative research. Evaluators or researchers hired to conduct theory-based or case-base analysis should be able to speak clearly about the difference between probability samples (simple random, systematic random, stratified, multi-stage cluster) and non-probability samples (convenience, snowball, quota, theoretical). They should be able to articulate to DfID staff and implementing partners how the different sampling strategies will affect the study including validity of findings and budget.

- *Research methodology*. Similarly, those hired should be able to explain to DfID staff and implementing partners the type of case-study, such as explanatory, instrumental, or exploratory, that is being conducted and why and articulate the steps they are taking to mitigate bias and ensure rigour. These might include:

  o Rich and detailed description that explores how underlying conditions interact with the intervention to produce results, allowing evaluators to identify some of the factors that may differentiate groups participating in and responding to the intervention from those who do not
  o Use of multiple data sources and triangulation of sources and methods (e.g., focus groups, interviews, observation, statistics)
  o Robust case study selection strategies
  o Triangulation of data sources and methods for collecting data (e.g., qualitative, quantitative, key informant interviews, focus groups, surveys, etc.), and

---

[67] See Maxwell, J. (2004) 'Using Qualitative Methods for Causal Explanation.' *Field Methods*. Vol. 16, No. 3: 243-264.
[68] White, H. (2009) *Some Reflections on Current Debates in Impact Evaluation*. 3ie Working Paper 1. Washington, DC: 3ie. p. 13.

o Robust strategies for ruling out alternative explanations described above (*see* section 6.1.5), such as General Elimination Method , searching for discrepant evidence and negative cases, and seeking feedback from others to 'check' the validity of the findings, surface any biases, assumptions or gaps in evidence[69].

The credibility of the evidence will depend on the quality of the evaluation, and it will be important for commissioners of impact evaluations using these approaches to explore the degree to which the various issues of bias have been addressed.

6.5 What are the conflict sensitivity concerns? Can they be addressed?

Ethics and Conflict Sensitivity are considerations that are just as important in theory-based and case based designs as they are in experimental methodologies. Evaluators have to ensure that when evaluating impact they are aware of how asking questions related to conflict may re-traumatise participants or trigger negative emotions. Given the prominence of qualitative methods in theory-based and case-based approaches, special attention should be given to the composition of the evaluation team, ensuring that the team itself possesses deep contextual knowledge can access stakeholders on different sides of the conflict divide. In addition, the potential impacts of data collection methods and the potential for the process and results to be perceived as biased or to exclude significant perspectives on the conflict are great. Qualitative methods may also require evaluators to be highly skilled in listening, culturally sensitive and skilled in facilitation and conflict resolution, especially if focus groups or other processes that bring people together are employed. In inception reports, DfID staff and implementing partners should see how evaluators aim to address ethical considerations and make evaluations conflict sensitive.

7   Participatory Approaches

7.1   What are they?

Participatory approaches to impact evaluation understand participants stakeholders as agents rather than passive actors. In participatory approaches, participants and other stakeholders are at the heart of data collection and analysis. They are engaged in many or all phases of the evaluation—from defining the questions to be answered, identifying and prioritising impact indicators, to deciding on methods, implementing the evaluation, interpreting the data and verifying results. Participatory approaches often do not have causal explanation as a primary focus, but rather accountability to stakeholders and joint learning to improve intervention strategies, implementation and sustainability of outcomes. However, they can and do collect evidence on causal linkages between interventions and impacts—often based on the actions and experiences and perceptions of stakeholders.

There is no single definition or methodology of participatory impact evaluation, but the designs share key principles and elements:

- Decisions about what impacts are the most relevant are made through dialogue with stakeholders. This allows the methods and focus of the evaluation to be most responsive to local needs and priorities.

---

[69] Maxwell, J. (2004) 'Using Qualitative Methods for Causal Explanation". *Field Methods*. Vol. 16,No. 3, pp. 243-64.

- Stakeholders are participants in the identification and interpretation of data. The key claim of participatory approaches is that causal and attribution questions are more reliably and accurately addressed through insights from stakeholders[70].
- Evaluations have reflection and action-orientation built in, to create learning and knowledge for stakeholders, programme implementers, donors and others to reflect on progress and take corrective action when needed.

Box 17: Participatory approaches as mechanisms for programme improvement

'A good [participatory] impact assessment process can create space for dialogue, and the results can provide a basis for discussions on how to improve programming and where best to allocate future resources.'

Source: Catley, A., Burns, J., Abebe, D. & Suji, O. (2013) *Participatory Impact Assessment: A Design Guide*. Somerville, MA: Feinstein International Center, Tufts University, p. 4.

Two approaches that have been used widely in evaluations of peacebuilding interventions include Most Significant Change (MSC) and Outcome Mapping/Outcome Harvesting. Participatory Impact Assessment (PIA) has been used increasingly in development and humanitarian evaluations, and has promise to be of value for impact evaluations in peacebuilding. These are described below.

7.5.1   Most Significant Change (MSC)

MSC is a form of outcome monitoring that focuses on collecting and selecting stakeholder accounts of significant changes that have occurred and deciding which is the most significant and why. The approach emphasises the need to respect stakeholders' own judgments regarding changes an initiative has made to their lives or in their community, although the approach includes validation of the stories provided through triangulation with other data sources. It is an inductive, goal-free method—with no pre-determined notion of what impacts ought to have been achieved, but rather eliciting stories (both positive and negative) of what impacts have been from stakeholders themselves. The central element of the approach is the systematic collection and selection of a purposive sample of significant change stories. Stories are collected from people most directly involved with an intervention (such as participants, field staff, affected community members) by asking a simple, open-ended question: 'what was the most significant change that took place for your community' (in a particular domain, such as relationships among people, over a particular period of time)? The most significant of the stories are selected through a multi-layered group process of review, selection—often (although not always) involving participants and community stakeholders.

MSC is a useful methodology for peacebuilding because it bases its judgments and analysis on the personal experiences of individual local stakeholders, helping programme implementers to understand in participant's own words how they view their environment and the change processes that take place within it. In a complex conflict context, MSC also can help to shed light onto which programme areas are most relevant to peace. MSC is generally not intended to be used as a stand-alone methodology for evaluating impact, but combined with other methods to investigate the reported changes and the causal links between the intervention and those impacts more deductively. In particular, a meta-analysis can also be used to identify trends that are crosscutting among a number of stories that should be subject to more thorough investigation.

---

[70] Menkhaus, K. (2004) *Impact Assessment in Post-Conflict Peacebuilding: Challenges and Future Directions*. Geneva: Interpeace. p. 11.

7.5.2   Outcome Harvesting

Outcome Harvesting is an evaluation design focused on 'harvesting' changes in behaviour, relationships, actions, activities, policies or practices of key actors for change. It is inspired by *Outcome mapping*, which was developed at the International Development Research Centre (IDRC) in Canada, as a participatory approach for planning, monitoring and evaluation of interventions. The distinguishing feature of outcome mapping and outcome harvesting is that they focus on how an intervention has contributed to changes in *behaviour* of and *relationships* among actors that it intends to influence and who are key to bringing about the macro-level impacts it intends to contribute to[71].

Outcome harvesting does not measure progress towards predetermined outcomes, but works with 'change agents' (intervention staff and boundary partners) to collect evidence of what has been achieved ('harvest' outcomes) and work backwards from these outcomes to document the who changed what (or whom), when and where, and how the intervention contributed to that change. For that reason has been found to be useful in complex situations where outcomes are unpredictable or changing, such as post-conflict peacebuilding environments.

Outcome harvesting identifies useable questions then collects information on outcomes from documentation to develop 'outcome descriptions' documenting the change in the key actors in specific and measurable ways. The 'outcome harvesters' (evaluators) engage with the change agents to review information, collect additional information and examine and substantiate the coherence, specificity and plausibility of the explanations of how the intervention has contributed to outcomes. Triangulation of data sources, including collection of outside views from people familiar with the intervention but independent from it, is used, and rigorous approaches such as General Elimination Method (see section 6.1.5) are used to rule out alternative explanations for the outcomes.

From a peacebuilding perspective, Outcome Harvesting can be useful because it identifies changes in behaviours and relationships that constitute intermediate steps (and actions) to realising broad peacebuilding outcomes such as tolerance, improved governance, reconciliation, improved rule of law, etc.; they identify specific program outcomes that might then be connected to impacts observed at the broader level of the conflict. However, in order for both of these methods to be truly participatory, it is especially important for participants to play a role in this process, particularly in unpacking the results chain between impact and the programme activities.

7.1.3 Participatory Impact Assessment (PIA)

Participatory Impact Assessment (PIA) is based on the methods and principles of Participatory Rural Appraisal (PRA), and like many other participatory methods, focuses on identifying impact for participants and the role of the programme in achieving those impacts. A PIA aims to understand the relative importance of programme activities against other events, interventions or processes that occurred independently of the programme. Participatory methods are used to a) identify the main factors that have led to changes in people's lives; b) categorises these factors as 'programme' or 'non-programme'; c) measure the relative importance of each. PIA uses an eight-step process that involves

---

[71] Outcome mapping has three broad stages: 1) planning ('intentional design') when the macro level changes to which the intervention wants to contribute are identified by the participants along with appropriate strategies and 'progress markers'—a set of graduated indicators of changed behaviours of key actors that represent incremental changes the programme realistically hopes to influence—to track performance; (2) outcome and performance monitoring using outcome journals, strategy journals and performance journals to collect and reflect on data related to the progress markers. (3) evaluation planning, may include outcome harvesting, where key outcomes are collected and measured against progress markers (if available). Although evaluation comprises the third stage, in practice, outcome mapping has focused primarily on planning and monitoring and thus may be of less use for impact evaluation.

working with participants in defining the evaluation scope and focus, selecting indicators and methods, analysis and triangulation of information, and verifying the results with the community. Many of the methods used in PIA are ranking or scoring methods, but are often followed by more qualitative in formal and open-ended interviews. Often, comparison methods such as before and after comparison, comparing different forms of programme support, and comparing participant groups are used to strengthen the rigour of PIA findings.

While PIA so far been used primarily in the humanitarian and development fields, for example to measure changes in people's livelihoods over time, it holds potential to help assess changes in the causes, symptoms and consequences of conflict as experienced by people and to understand how peacebuilding programmes and other factors contributed to them.

7.2   What kind of peacebuilding impact evaluation questions is this approach useful for?

Participatory approaches offer a unique opportunity to gain a close understanding of the perspective of stakeholders, to understand context-specific impacts and causal chains in detail, and to align evaluation questions and methods with the priorities of stakeholders, therefore increasing ownership.

With this in mind, participatory approaches have greatest utility in:

- *Questions for which effects on stakeholders and stakeholders' views of progress are particularly important.* In conflict situations, where competing visions of what a 'peaceful' society should look like, differing interests, needs and grievances, and negative experiences of the 'other' are likely to affect how progress is defined and perceived by different groups, understanding how interventions affect different groups differentially and are interpreted by them differentially can be important. Stakeholders' views of desirable goals and of 'success' may also be different from those of donors and implementers. When this is the case, participatory approaches can develop more nuanced understandings of what outcomes that are valued and different ways in which interventions are experienced by different groups.

- *Improving interventions.* When the purpose of the evaluation is to learn about what works and does not in order to take decisions about whether or not to continue and how to adjust or follow up to a programme, participatory approaches can be very helpful. Like theory-based and case-based approaches, with which they are often integrated, participatory approaches identify *how and why an intervention cause impacts*, especially how contextual factors, programme implementation and embedded assumptions about change may have affected impact. This provides valuable and actionable knowledge for redesign and adjustment. In addition, because participatory approaches strengthen *ownership* of the initiative and of the evaluation process itself, the evaluation results are likely to be more useful and participants more motivated to contribute ideas on how to improve the intervention and apply lessons learned.

- *Evaluating sustained outcomes.* Participatory methods are designed to encourage stakeholder ownership of the intervention and its results, as well as develop skills and capacities to monitor and evaluate among stakeholders. This creates a foundation for undertaking ongoing evaluation beyond the life of the programme itself. Moreover, stakeholders' knowledge and understanding of the context and their own circumstances can help to identify long-term changes and untangle the reasons they have been sustained—including whether and how they might be due to the intervention.

- *Evaluating unplanned and unanticipated results.* The engagement of stakeholders in defining relevant 'impacts' and exploring them allows for identification of impacts that may not have been part of an intervention's plan or theory of change. They are particularly relevant for evaluating unintentional impacts on conflict and impacts of innovative interventions where the theory of change may not be well-developed. For the latter, participatory approaches may be able to develop a theory of change empirically—by identifying ways in which the intervention contributed to observed impacts.

Box 18: Integrating participatory approaches in a study of cumulative impacts of peacebuilding in Kosovo

In a study on the impact of peacebuilding interventions on certain communities' non-participation in violence in Kosovo in 2004 (see Box 6), participatory methods were used at the outset to define the evaluation questions and issues of interest and identify relevant case studies. This included defining violence and how the context of violence had evolved prior to 2004. Consultations with stakeholders were integrated into the process to analyse the evidence gathered (both quantitative and case study). These provided important insights into how data should be interpreted, the mechanisms by which security improved, and how different the ways Albanians and Serbs experienced and viewed security and sense of security were. Although 'objective' indicators of violence (physical harm) clearly had improved, the other ways in which 'security' had deteriorated were uncovered and ways in which perpetrators of violence had adapted to the presence of international peacekeepers to instil fear and insecurity uncovered. As a result, the study developed more nuanced and actionable analysis of the ways peacebuilding and peacekeeping interventions affected security, and developed a deeper understanding of the significance of non-participation in the 2004 violence.

Source: CDA (2006) *Has Peacebuilding Made a Difference in Kosovo? A Study of the Effectiveness of Peacebuilding in Preventing Violence—Lessons Learned from the March 2004 Riots in Kosovo.* Cambridge, MA and Pristina, Kosovo: CDA and CARE International.

7.3 What kinds of peacebuilding impact evaluation questions are these approaches not useful for?

There are important limitations to what questions participatory methods can answer. Due to the close focus on stakeholder involvement, participatory methods used alone offer less explanatory power for changes at the macro conflict level ('peace writ large') or across a large number of geographic locations. Where participants or stakeholders for macro-level change are hard to define or too numerous, or for large, multi-level, multi-site comprehensive evaluations, participatory approaches are difficult to implement—both due to time and cost considerations, but also conceptual consideration about who ought to be included and not in the process. For these reasons, participatory approaches are often combined with other methods.

In addition, where the intervention addresses a polarised political or social situation—such as a mediation process, or track two dialogue between different ethnic groups—participatory approaches may be difficult and inappropriate, as participatory evaluation is essentially a 'social process for negotiating between people's different needs, expectations and world-views'[72]. In polarised conflict situations, where the intervention itself represents an effort to bridge people's different (and often opposing) views, aspirations and perspectives, it could be difficult for participants to reach consensus on a common set of indicators or outcomes; a participatory approach may be inappropriate and potentially conflict insensitive, even if solicitation of the different views and perceptions of stakeholders from different groups would be a necessary source of evidence.

---

[72] Estrella, M. & Gaventa, J. (1998) *Who Counts Reality? Participatory Monitoring and Evaluation: A Literature Review*. IDS Working Paper 70. https://www.ids.ac.uk/files/Wp70.pdf.

7.4   What are methodological strengths of these approaches in evaluations of peacebuilding impact?

Participatory approaches offer a number of methodological advantages that address gaps in other approaches:

- *'Construct validity.'*   Participant engagement in the definition of the evaluation questions and the measures to be used enhance the likelihood that the measures of impacts are real, meaningful reflections of what is happening on the ground. In the study of the impacts of peacebuilding on violence in Kosovo (Boxes 6 and 18 above), for example, integration of participatory approaches into the study at the stage of definition of the questions to be investigated and the definition of the outcomes ('violence' and 'lack of violence') led to a more multi-dimensional and nuanced conception and measure of the 'impacts' to which peacebuilding interventions were to have contributed.

- *Better data collection.* Done well, participatory approaches can promote trust and confidence, which facilitates openness and honesty about information. With stakeholder partnership, evaluators may be greater access to key informants and constituencies or voices that may be invisible, gain access to local knowledge and information that people may not be willing to reveal to 'outsiders' (even national 'outsiders'); participatory approaches may surface issues in the communities or stakeholders that an outsider may not have been aware of but that may significantly affect the evaluation findings. This can be particularly important on highly sensitive topics and highly politicised environments.

- *Better interpretation of results.*   It can be difficult to interpret meaning of the data without nuanced contextual knowledge. Stakeholder participation in analysis and interpretation of the evidence can lead to more accurate and more nuanced evaluation conclusions.

- *Minimisation of evaluator bias.*   Biases of evaluators—ranging from implicit or explicit assumptions about conflict dynamics and how peace is built to what voices are important to include in assessing impacts—are subject to discussion and challenge. This helps ensure that the evaluation and the findings do not reflect the evaluator's biases and miss evidence that is not consistent with their assumptions about the conflict or theories about how conflicts transform.

7.5   What are methodological challenges and threats to data validity of these approaches for evaluation of peacebuilding impacts?

- *Rigour about causality.*   The focus of participatory approaches is not primarily on attribution or contribution; their goal is to collect actionable knowledge and understandings to improve programming. If not implemented rigorously, they can rely heavily on participant accounts of contribution, which may make the evaluation subjective, anecdotal, and uncritical, leading to incomplete assessment of impact. Data collection by participants can at times (depending on the level of skill) also be less rigorous or technically flawed. Therefore, in order to deliver rigorous

findings, those conducting impact evaluations need to invest in the skills and capacities of local partners, triangulate data through multiple sources, and look for robust validation and specification of stories and evidence. To further increase rigour, participatory approaches are also often combined with or integrated into other methods.

- *Bias.* In conflict contexts, a danger of 'groupthink' or a common narrative about the conflict or about certain types of interventions may bias results. Participatory approaches requires some dependence on recollections of the intervention, which can be faulty or distorted, although if undertaken at the design stage of an intervention is likely to create documentation and continuing awareness of the efforts to mitigate that.

- *Who participates?* In all impact evaluations bias and subjectivity of results will depend to a large extent in these methods on who participates in the evaluation and whose voices are heard. For participatory approaches, the question of who participates poses a greater risk. On the one hand, because participation consists not only of solicitation of stakeholder perceptions, but involves actual engagement in multiple stages of the evaluation process and there is a risk of introducing conflict-related biases at multiple points of the evaluation, especially if important voices are excluded for some reason. On the other hand, with the potential exception of Participatory Impact Assessment, these methods are not *inherently* participatory; the degree to which they are truly 'participatory', depends greatly on how they are implemented, in particular the degree and frequency to which stakeholders are involved in designing the evaluation focus and process, collecting the relevant data, interpreting the data, and identifying the significance of the results based on their own priorities and understanding of the local context. This issue takes on special importance in conflict situations because they are characterised by vastly different, and often polarised views and perceptions about what is happening and why, and there is a risk of adopting one party's point of view or bias. Ensuring that key stakeholder groups engage represent important (different) perspectives and experiences in relation to the conflict context will be important in ensuring that results do not reflect one particular view or understanding of the intervention, and to avoid negative conflict sensitivity consequences.

7.6   What are the conflict sensitivity concerns? Can they be addressed?

Participatory approaches potentially communicate important, positive *implicit ethical messages*[73] that can affect the conflict context positively. They promote consistency with peacebuilding and development values regarding accountability to beneficiaries and ownership, respect, transparency and partnership. These can have important signalling effects that can affect the perceptions of the intervention and its implementers, as well as relationships among participants and the government themselves.

However, like other methods, they risk worsening conflict-related cleavages, in particular if perspectives on the goals and implementation of an intervention, what impacts are significant or even what impacts can be observed, as well as the interpretation of the evidence are different and contentious. While the participatory process may provide an opportunity for deep dialogue on these questions, it should be recognised that where groups are polarised, or otherwise have difficulty working together, collaboration needed for participatory evaluation to work well may not be possible—or may require special facilitation skills on the part of the outside evaluator or programme staff to succeed.

---

[73] Anderson, M. (1999) *Do No Harm: How Aid Can Support Peace—Or War*. Boulder, CO: Lynne Rienner.

Annexes

Annex 1: Strengths and Limitations of Different Approaches to Designing Impact Evaluations of Peacebuilding Interventions

Experimental and quasi-experimental, mechanism-based, and participatory methodologies offer a variety of strengths and weaknesses when designing an impact evaluation. The following table summarises these attributes. Validity refers to the accuracy and rigour of measurement—how well the methodology measures what it intends to measure. Utility refers to the types of evaluation questions each approach may be useful in answering.

Approach	Validity: Accuracy of measurement		Utility: What sorts of evaluation questions is the approach useful for?	
	Strengths	**Limitations**	**Strengths**	**Limitations**

Experimental and quasi-experimental	When requirements for valid design can be met:
	• Establishes a counterfactual or valid comparison, reducing bias of overestimating or underestimating impact
	• Attribution/contribution quantified—provides evidence on how much the intervention contributed
	• Causal clarification is clear, transparent
	• Rigorous reduction of selection bias—distortion of results when participants in a program are systematically different from non-participants (even before they enter the program).
	• Often carried out in real time, ensuring there is baseline data and on-going tailored data on the question from the beginning of the intervention
		• Strict requirements for valid design are difficult to meet in fragile and conflict-affected states, e.g.:
		  o Difficult to establish comparison and control groups and 'control' them—randomisation may not be random
		  o Difficult to keep intervention constant throughout time period of intervention and across locations
		  o Risk of high drop-out rates from treatment or control groups can undermine statistical significance
		• Weak ability to determine causation in complex environments, including multi-variable causality or non-linear results chains
		• Risk of inaccurate findings due to timing of when impacts may be visible, especially if after the end of the intervention
		• Construct validity: it may be difficult to ensure that the
			• Assessing whether interventions caused a particular impact on participants, and, if so, by how much
			• Testing specific links in the chain of results of a theory of change
			• Validation of proof of concept—assessing viability of a strategy or approach in different circumstances
			• Comparing alternative intervention strategies or activities in a specific context
				• Does not provide answers regarding:
				  o Effects on 'peace writ large' or drivers of conflict
				  o Higher-level outcomes such as policy, political and legislative processes, behaviour or a single institution, culture change, mediation
				  o Unintended or unanticipated impacts
				• Cannot assess HOW or WHY an intervention achieved impact
				• Not suitable for decisions about improving interventions or replicating them elsewhere
				• Difficult to assess sustained or long term impacts on "peace writ large" or conflict drivers
				• Does not differentiate effects of interventions (who benefits/not, where it works well and where it works less well)

		• evaluation methods actually measure the seemingly ambiguous objectives most common in peacebuilding (e.g. social cohesion, inclusive governance) • Risk of misinterpretation of findings due to decontextualisation and lack of understanding of how impacts occurred • Use of control groups and behavioural games may exacerbate conflict dynamics and drivers		
Mechanism-based (theory and case-based)	• Realistic assessment of causality in complex, multi-dimensional settings • Rigor through triangulation translates into strong findings about causation (but not as strong as experimental) • Intensive, contextual, holistic understanding of a specific intervention • Construct validity: the evaluation methods are likely to actually measure the seemingly ambiguous objectives most common in peacebuilding (e.g. social cohesion, inclusive governance) • External validity: evaluation results can be generalised and extended to other situations	• Risk of selection bias: Program participants may be different from non-participants based on sampling method, or bias in case selection • Risk of correlation bias: narrative emphasis in methods and analysis may lead to confusion between causation and correlation. There are many strategies for mitigating this bias, though often they are not employed. • History of non-rigorous implementation, especially relating to sampling and mitigation of bias; while not inherent in methodology, risks are high due to poor quality/implementation • May require high facilitation and conflict resolution skills for evaluators in data collection	• Good for answering questions related to: ○ Program contribution to higher level impacts, e.g., policy, culture, and institutional change ○ Effects on 'peace writ large' and drivers of conflict ○ Unintended and unanticipated impacts ○ Conflict sensitivity of development, humanitarian and peacebuilding interventions ○ Sustained impacts • Helps explain HOW and WHY an intervention caused impact • Tests implicit and explicit assumptions in theory of change	• Difficult to obtain precise estimates of attribution
Participatory	• Better able to integrate views of project participants • Provides a realistic assessment	• Risk of limited rigour about causality: participatory approaches focus on actionable	• Good for answering questions regarding: ○ Individual-level change	• Limited ability to quantify impact (although there are quantitative participatory methods)

	of causality in complex, multi-dimensional settings • Can deal with complexity • Construct validity: the evaluation methods are likely to actually measure the seemingly ambiguous objectives most common in peacebuilding, such as social cohesion and cohesive governance • Better data collection, especially on sensitive topics. • More nuanced and contextual interpretation of results • Helps minimise evaluator bias by subjecting results to scrutiny from participants	knowledge to improve programming over attribution or contribution and data collection relies heavily on local accounts of events and local data collection efforts—rigorous triangulation and validation must be built into evaluation design • Risk of bias, including group think, altered recollections of events. • Risk of introducing conflict-related biases due to exclusion of voices—ensuring participants represent important (different) perspectives and experiences in relation to conflict can be difficult • Can risk deepening conflict-related cleavages if perspectives on goals, implementation, what impacts are significant are differences, or if interpretation of findings is contentious	o Impacts related to or for which understanding of effects on stakeholders and stakeholders' views of progress are particularly important o How interventions affect different groups differently and are viewed differently o Unexpected outcomes or impacts o HOW and WHY an intervention caused impact o Sustained impacts • Useful where purpose is to improve intervention • Evaluation use enhanced through participant ownership over evaluation process and results • Improving interventions through participant ownership over evaluation process and results • Consistency with peacebuilding values of participation, inclusiveness, and accountability and joint learning with participants	• May be of limited utility or difficult to implement in highly contested situations or issues • May have less explanatory power for changes at macro-level or across a large number of geographic locations, unless combined with other methods

Annex 2: List of Resources

General Resources

Andersen, O., Bull, B. & Kennedy-Chouane, M. Eds. (2014) *Evaluation Methodologies for Aid in Conflict*. London: Routledge.

Gaarder, M. & Annan, J. (2013) *Impact Evaluation of Conflict Prevention and Peacebuilding Interventions*. Policy Research Working Paper 6496. Washington, D.C.: World Bank. http://elibrary.worldbank.org/doi/pdf/10.1596/1813-9450-6496.

Interaction (2012) Interaction Evaluation Note and Webinar Series http://www.interaction.org/impact-evaluation-notes. Includes four guidance notes and webinar recordings and slides on each.
- Rogers, P. (2012) *Introduction to Impact Evaluation*. Impact Evaluation Notes No. 1.
- Perrin, B. (2012) *Linking Monitoring and Evaluation to Impact Evaluation*. Impact Evaluation Notes No. 2.
- Bamberger, M. (2012) *Introduction to Mixed Methods in Impact Evaluation*. Impact Evaluation Notes No. 3
- Bonbright, D. (2012) *Use of Impact Evaluation Results*. Impact Evaluation Notes No. 4.Leeuw, F. & Vaessen, J. (2009) *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation*. Washington, D.C.: Network of Networks for Impact Evaluation (NONIE)/Independent Evaluation Group, World Bank. http://siteresources.worldbank.org/EXTOED/Resources/nonie_guidance.pdf.

Rogers, M. (2012) *Evaluating Impact in Conflict Prevention and Peacebuilding Programs*. Cambridge, MA: CDA. http://cdacollaborative.org/media/123704/Evaluating-Impact-in-Conflict-Prevention-and-Peacebuilding-Programs.pdf.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. & Befani, B. (2012) *Broadening the Range of Designs and Methods for Impact Evaluations: Report of a study commissioned by the Department for International Development*. DFID Working Paper 38. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf.

White, H. (2009) *Some Reflections on Current Debates in Impact Evaluation*. International Initiative for Impact Evaluation. Washington, D.C.: 3ie.

Experimental, Quasi-Experimental and Statistical Approaches

Gertler, P., Martinez, S., Premand, P., Rawlings, L. & Vermeersch, C. (2011) *Impact Evaluation in Practice*. Washington, D.C.: World Bank. http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1295455628620/Impact_Evaluation_in_Practice.pdf.

Glennerster, R. & Takavarasha, K. (2013) *Running Randomized Evaluations: A Practical Guide*. Princeton: Princeton University Press.

Mechanism-Based Approaches (Theory-based and case-based)

General Resources and Overviews

European Commission (no date) *Evalsed Sourcebook: Method and Techniques*. Brussels: European Commission. http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf.

Leeuw, F. (2012) *Theory-Based Evaluation*. Brussels: European Commission. http://www.alnap.org/node/13651.aspx.

White, H. and Phillips, D. (2012) *Addressing attribution of cause and effect in small n impact evaluations: toward an integrated framework.* 3ie Working Paper 15. New Delhi: International Initiative for

Impact Evaluation.
http://www.3ieimpact.org/media/filer_public/2012/06/29/working_paper_15.pdf.

White, H. (2009) *Theory-Based Impact Evaluation: Principles and Practice*. International Initiative for Impact Evaluation Working Paper 3. Washington, D.C.: 3ie.
http://www.3ieimpact.org/media/filer_public/2012/05/07/Working_Paper_3.pdf.

Realist Evaluation

Pawson, R. & Tilley, N. (1997) *Realistic Evaluation*. London: Sage.

Pawson, R. & Tilley, N. (2004) 'Realist Evaluation'. Paper.
http://www.communitymatters.com.au/RE_chapter.pdf.

Realist Evaluation. Description and resources at
http://betterevaluation.org/approach/realist_evaluation.

Contribution analysis

Mayne, J. (2011) 'Contribution Analysis: Addressing Cause and Effect'. In K. Forss, K., Marra, M., and Schwartz, R. (eds.) *Evaluating the Complex*. New Brunswick, NJ: Transaction Publishers.

Mayne, J. (2008) 'Contribution Analysis: An approach to exploring cause and effect'. ILAC Brief 16. The Institutional Learning and Change (ILAC) Initiative, Bioversity International. http://www.cgiar-ilac.org/files/publications/briefs/ILAC_Brief16_Contribution_Analysis.pdf.

Mayne, J. (2001) 'Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly'. *Canadian Journal of Program Evaluation*. Vol. 16, No. 1: 1-24.

Contribution Analysis. Description and Resources at
http://betterevaluation.org/plan/approach/contribution_analysis.

Process Tracing

Bennett, A. (2010) 'Process tracing and causal inference' In Brady, H. and Collier, D. (eds.) *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman and Littlefield.

Collier, D. (2011) 'Understanding process tracing'. *Political Science and Politics*. Vol. 44(4): 823-30.

George, A. L. and Bennett, A. (2005) *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.

Oxfam (no date) *Process Tracing: Draft Protocol*. http://policy-practice.oxfam.org.uk/~/media/Files/policy_and_practice/methods_approaches/effectiveness/Process-tracing-draft-protocol-110113.ashx.

Case studies

US General Accounting Office, Program Evaluation and Methodology Division (GAO) (1990) *Case Study Evaluations*. GAO/PEMO-91-10.1.9. Washington, D.C.: GAO.

Neale, P., Thapa, S. & Boyce, C. (2006) *Preparing a Case Study: A Guide for Designing and Conducting a Case Study for Evaluation Input*. Watertown, MA: Pathfinder International.
http://inalytics.flance.co.uk/wp-content/uploads/2013/09/casestudy.pdf.

Simons, H. (2009). *Case study research in practice*. London: SAGE Publications Ltd.

Woolcock, M. (2013) 'Using case studies to explore the external validity of "complex" development interventions'. WIDER Working Paper No. 2013/096. UNU-WIDER.
http://www.wider.unu.edu/publications/working-papers/2013/en_GB/wp2013-096/.

General elimination method

Patton, M. Q. (2008) 'Advocacy impact evaluation'. *Journal of Multi-Disciplinary Evaluation* 5 (9), March 2008.

Scriven, M. (1976) 'Maximising the power of causal investigations: the Modus Operandi method'. In Gene V. Glass (ed.) *Evaluation Studies Review Annual, Vol. 1*. London: Sage Publications.

Participatory Approaches

Most Significant Change

Davies, R. & Dart, J. (2005) *The 'Most Significant Change' (MSC) Technique: A Guide to Its Use*. http://www.mande.co.uk/docs/MSCGuide.pdf.

Most Significant Change. Description and resources at http://betterevaluation.org/plan/approach/most_significant_change.

Outcome mapping and Outcome harvesting

Smutylo, T. (2005) *Outcome mapping: A method for tracking behavioural changes in development programs*. ILAC Brief 7. http://www.outcomemapping.ca/resource/resource.php?id=182.

Earl, S. (2008) Introduction to Outcome Mapping. Three videos providing an introduction to the concepts of outcome mapping. https://www.youtube.com/watch?v=fPL_KEUawnc.

Wilson-Grau, R. & Britt, H. (2013) *Outcome Harvesting* (revised). Cairo: Ford Foundation. http://www.outcomemapping.ca/resource/resource.php?id=374.

World Bank Institute (2014) *Cases in Outcome Harvesting: Ten pilot experiences identify new learning from multi-stakeholder projects to improve results*. Washington, D.C.: World Bank. http://wbi.worldbank.org/wbi/Data/wbi/wbicms/files/drupal-acquia/wbi/Cases%20in%20Outcome%20Harvesting.pdf.

Outcome mapping. Description and resources at http://betterevaluation.org/plan/approach/outcome_mapping.

Participatory Impact Assessment

Catley, A., Burns, J., Abebe, D., Suji, O. (2014) *Participatory Impact Assessment: A Design Guide*. Medford, MA: Feinstein International Center, Tufts University. http://fic.tufts.edu/publication-item/participatory-impact-assessment-a-design-guide/.

Estrella, M. & Gaventa, J. (1998) *Who Counts Reality? Participatory Monitoring and Evaluation: A Literature Review*. IDS Working Paper 70. Sussex: Institute of Development Studies. https://www.ids.ac.uk/files/Wp70.pdf.

Estrella, M., Blauert, J., Campilan, D., Gaventa, J., Gonsalves, J., Guijt, I., Johnson, D. & Ricafort, R. (2000) *Learning from Change: Issues and Experiences in Participatory Monitoring and Evaluation*. London and Ottawa: Intermediate Technology Publications and IDRC. http://www.idrc.ca/EN/Resources/Publications/Pages/IDRCBookDetails.aspx?PublicationID=348.

Participatory evaluation. Description and resources at http://betterevaluation.org/plan/approach/participatory_evaluation.