Measuring the Impact of Peacebuilding Interventions on Rule of Law and Security Institutions

Vincenza Scherrer

DCAF

a centre for security,
development and
the rule of law

Measuring the Impact of Peacebuilding Interventions on Rule of Law and Security Institutions

Vincenza Scherrer

───────────

DCAF

This book was originally published by the Geneva Centre for the Democratic Control of Armed Forces (DCAF), an international foundation whose mission is to assist the international community in pursuing good governance and reform of the security sector. The title transferred to Ubiquity Press when the series moved to an open access platform. The full text of this book was peer reviewed according to the original publisher's policy at the time. The original ISBN for this title was 978-92-9222-204-8.

SSR Papers is a flagship DCAF publication series intended to contribute innovative thinking on important themes and approaches relating to security sector reform (SSR) in the broader context of security sector governance (SSG). Papers provide original and provocative analysis on topics that are directly linked to the challenges of a governance-driven security sector reform agenda. SSR Papers are intended for researchers, policy-makers and practitioners involved in this field.

The views expressed are those of the author(s) alone and do not in any way reflect the views of the institutions referred to or represented within this paper.

Contents

INTRODUCTION[1]

Peacebuilding aims to lay the foundations for sustainable peace and development by promoting measures that seek to reduce the risk of violent conflict.[2] Since the 1990s, internationally-supported peacebuilding interventions have become increasingly prominent. The number of international peacebuilding interventions, however, stands in stark contrast to our limited knowledge of their impact on the ground. Better understanding the impact of these interventions is thus a prerequisite for improving peacebuilding practice. In short, this means asking what is working, what is not, and why? At the same time today's difficult economic climate has only added impetus to the need to understand the impact of peacebuilding interventions. Yet if the need to measure impact appears self-evident, conceptual advances in how to measure it nevertheless remain limited. This is simply unsustainable given the need to maximise the effectiveness of what are major international commitments.

Activities focusing on rule of law and security institutions are a key component of the peacebuilding agenda. According to the Capstone doctrine for UN peacekeeping operations, 'Security Sector Reform (SSR) and other rule of law-related activities' constitute 'critical peacebuilding activities'.[3] In simple terms, rule of law and security institutions are responsible for the provision, management and oversight of justice and security in a country. Legal, judicial, prison and police institutions constitute core rule of law institutions.[4] Security institutions can include a wide variety of actors, such as the armed forces, police, corrections, intelligence services

and institutions responsible for border management, customs and civil emergencies, as well as those bodies that manage and oversee the delivery of security such as ministries, legislative bodies and civil society groups.[5] Drawing on a background study prepared for the Office of Rule of Law and Security Institutions (OROLSI) at the United Nations Department of Peacekeeping Operations (DPKO), [6] this paper purposefully uses the terminology of OROLSI, which links rule of law and security institutions to areas such as justice, corrections, police, disarmament, demobilization and reintegration, security sector reform and mine action.[7]

While policy frameworks concur that activities should be interrelated, support to rule of law and security institutions tends to be delivered in a siloed fashion. Moreover, programming often takes place on the assumption that activities will lead to positive, long-term change that will be felt in the lives of beneficiaries. Rarely are these assumptions tested. In fact, evaluations – when they occur – are often heavily focused on 'outputs' that do not shed light on the value of international support. Impact measurement can therefore help to focus on the bigger picture, understanding how efforts complement or contradict one another.[8] Such an approach can offer major benefits for the coherence, sequencing, and coordination of international support.[9]

Pressure to measure the impact of international support to rule of law and security institutions in peacebuilding interventions will only increase.[10] Demands for a better understanding of impact measurement began in the development community and have now spread to the humanitarian and peacebuilding communities. These demands have sometimes met with resistance but they have also triggered confusion over when and how impact can be measured. Impact has often been perceived as a particularly elusive level of the results chain where the contribution of an intervention cannot be proven.[11] Furthermore, there is a tendency to perceive impact as being visible only several years after an intervention and therefore as too long term to be measured effectively for the purposes of programming and policy. The high costs of effective impact assessment are also put forward as a disincentive. However, if there is scepticism over the feasibility of determining impact, there is also a growing consensus that impact measurement can be 'demystified'.[12] This would result in much needed clarity on a number of levels: 'to know whether the intervention

has worked, to learn from it, to increase transparency of the intervention, and to know its "value for money"'.[13]

The objective of this paper is to better understand ways to measure the impact of peacebuilding interventions on rule of law and security institutions. In particular, it seeks to identify impact assessment methodologies that can be applied in complex, multi-layered post-conflict interventions. Emphasis is placed on how bilateral and multilateral actors have approached this challenge in practice, with a particular focus on the role of the United Nations.[14]

The measurement of impact poses a number of dilemmas, including weighing the relatively high costs of assessments against the potential benefits and disaggregating the contributions of specific interventions given the wide range of actors concerned. In view of these concerns, this paper highlights methodologies that can be used for measuring impact drawing on both qualitative and quantitative approaches. While it is sometimes assumed that impact can only be measured using scientific-experimental approaches based on control trials and statistical methods, there are a number of other applicable methodologies. This understanding is an essential starting point in broadening the range of methods and approaches available for impact measurement and thereby increasing the ability of international actors to apply the methodology best suited to the immediate context, skills and resources available. This is particularly important in the area of post-conflict peacebuilding, where the international community often struggles to meet its basic monitoring and evaluation (M&E) requirements due to capacity and resource gaps.

This study draws on a broad review of current evaluation approaches and methodologies based on extensive desk research. Narrowing the focus to the approaches most relevant for measuring impact, methodologies focussing substantially on measuring performance at the output level were discarded. The evaluation approaches of nineteen international actors (bilateral and multilateral) were reviewed to decipher how they approach the need to measure impact.[15] The research examines primary sources (e.g. evaluations, official documents, reviews) and secondary sources (e.g. handbooks, guidance notes). Meta-evaluations, which evaluate the evaluations of international actors, were found to be particularly useful because they compare the approaches used in dozens of evaluations providing detailed information on the average timing of evaluation projects,

the composition of evaluation teams and the challenges encountered. Finally, the policy and academic literature on evaluation in post-conflict contexts and impact assessments provided useful background to the analysis.[16] While this study cannot claim to cover all possible approaches and methodologies for measuring impact, it has identified those considered most useful for international actors seeking to measure the impact of peacebuilding support to rule of law and security institutions.[17]

This study is based on definitions found in the OECD DAC Glossary of Key Terms in Evaluation and Results Based Management, a reference commonly used by international actors. 'Impact' is defined as 'positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended'.[18] The definition of impact *stricto sensu* as contained in the OECD DAC glossary is that it is 'produced by a development intervention' – thus suggesting that impact can be 'attributed' to a specific intervention. However, when it comes to measuring impact there is a debate about the validity of *attribution* versus *contribution*. Attribution is often promoted as the 'gold standard' because of its ability to demonstrate a direct causal link between an intervention and its impact. However, in complex post-conflict settings it is considered extremely difficult to isolate the effects of a particular peacebuilding intervention and thus to establish a causal link between the intervention and the observed outcomes and impacts. To avoid this so-called 'attribution gap',[19] efforts have increased to demonstrate the plausible contribution of an intervention to observed outcomes and impacts. Focusing on contribution as opposed to attribution recognises that there may be other factors that have also contributed to the observed impact. This is particularly relevant in post-conflict contexts as it takes into account the complexity of 'tracking causality' in the 'non-linear multi-agency contexts' within which peacebuilding support takes place.[20]

This paper uses the term 'impact assessment' as opposed to 'impact evaluation' in order to avoid confusion with the scientific-experimental approach to evaluating impact which is commonly referred to as 'impact evaluation'. The term 'evaluation' is used in general terms while 'assessment' is used when following the word 'impact'. The definition of 'evaluation', based on OECD usage, is 'the systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation or results. The aim is to determine the relevance

and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability.'[21] Impact is thus only one of five distinct criteria that might be used in evaluations, depending on what the evaluation is seeking to understand.

'Impact assessment methodologies' are therefore understood as those methodologies that seek to evaluate impact (including the negative, positive, intended and unintended effects of an intervention). The study does not examine impact assessments that are conducted *ex ante* to assess the potential future effects of an intervention. Instead the focus is on those approaches that could be adapted to analyse support to rule of law and security institutions, and which could be of relevance to peacebuilding. While monitoring and evaluation (M&E) are closely linked, this paper focuses specifically on evaluations. Monitoring is thus addressed only in so far as it relates to the goal of evaluation.

The paper begins with an overview of the methodological approaches and methodologies that can be used to evaluate impact. The experience of international actors in promoting and measuring impact are then examined to identify how international actors are currently approaching this issue. The paper then outlines the building blocks of an approach to measuring the impact of peacebuilding interventions in support of rule of law and security institutions in host countries. The paper concludes by identifying key findings and recommendations.

IMPACT ASSESSMENT: APPROACHES AND METHODOLOGIES

This section provides a brief overview of the origins and development of impact assessments, highlighting their evolution from assessment of the potential future consequences of programmes to examining the effects of interventions. The methodological approaches – providing frameworks within which the specific methodologies can be understood – are presented and summarised. The discussion is intended to facilitate understanding of what tools are available depending on the desire of individual actors to use a specific evaluation approach.

Origins and Development of Impact Assessments

Impact assessments were originally developed and used in the public sector to assess the potential future effects – intended and unintended, negative and positive – of public policies (*ex ante* assessments).[22] This practice was then adopted by the development community, with international agencies seeking to mitigate any negative consequences of their activities by assessing the potential future environmental, social, health and economic impacts of their interventions.[23] With the increasing emphasis on aid effectiveness in the 1990s, impact assessments began to be used for post-intervention evaluation and not just as a pre-intervention planning tool.[24] Impact has thus become one of the main evaluation criteria in the OECD DAC guidelines (alongside relevance, efficiency, effectiveness and sustainability). There has also been increased interest in the creation of

methodologies to assess and measure impact – mainly with a view to assessing the long-term effects of an intervention.

A similar evolution can be seen in the field of peacebuilding. The term 'impact' was first used by Kenneth Bush's Peace and Conflict Impact Assessment (PCIA) approach in 1998.[25] This approach aims to assess *ex ante* the potential impacts of development projects within the wider context of peace and conflict. Mary Anderson's 'do no harm' approach (1996) is a similar *ex ante* assessment of the potential negative impacts of development and humanitarian interventions, and prepared the ground for the widely accepted principle of conflict sensitivity.[26] Thania Paffenholz and Luc Reychler developed the 'Aid for Peace' approach in 2007[27] – also seen as third-generation PCIA – which proposes a four-stage impact assessment (needs, relevance, risks and effects) for development, humanitarian and peacebuilding interventions.[28]

Interest in monitoring and evaluation of peacebuilding interventions has thus grown since the late 1990s and early 2000s.[29] Increasing demands for proven effectiveness and measurable results in the peacebuilding field has raised awareness of the need to measure impact, but at the same time measuring impact in complex post-conflict settings has been considered especially difficult.[30] There is a school of thought that suggests it is impossible to attribute impacts to a given peacebuilding intervention: in particular, it has been suggested that to measure impact a great deal of sophisticated data collection and analysis over a long period of time is needed, and that 'these requirements either exceed the capacity of many organisations practicing peacebuilding or they extend beyond the donors' funding period'.[31] There have therefore been calls to focus on outcomes rather than impact.[32] Other voices argue that it is possible to assess impact by relying exclusively on scientific-experimental approaches like those applied in development interventions.[33] But in practice this approach is seldom used in peacebuilding contexts due to its resource-intensive nature and the fact that it is most usefully applied *ex post*.[34] Yet as in the development field, other approaches to measuring impact that focus on 'contribution' as opposed to 'attribution' are attracting increasing recognition.[35]

One of the most important contributions to impact assessments in peacebuilding evaluations stems from Collaborative Learning Project's Reflecting on Peace Practice (RPP). RPP collaborators challenge the

resistance of many peacebuilding practitioners to assessing impacts. In particular, they point out that impact has to be 'demystified' and 'is not always elusive and unreachable, too long-term or impossible to assess'.[36] There can be short-term impacts and long-term impacts. The authors advocate that each peace project be held accountable for its contribution to the broader peace, or 'peace writ large'.[37] The RPP approach provided the basis for new OECD DAC guidance on Evaluation of Conflict Prevention and Peacebuilding Activities.[38] The guidance highlights the challenges of M&E in conflict and post-conflict settings, and combines the OECD DAC evaluation criteria (relevance, efficiency, impact, effectiveness, sustainability and coherence) with a conflict-sensitive M&E approach. Most importantly, it takes up the RPP definition of impact, noting that it can also look at short-term and not just long-term impact.[39]

In sum, while impact assessments stem from attempts to assess future consequences of activities, they have been adopted as a tool to measure the impact of ongoing or completed projects. Moreover, while traditional understandings of evaluation at the impact level was that they had to be undertaken *ex post* (several years after an intervention is in place or completed), there is a growing view that impact can be measured in the more immediate term. This emerging approach offers opportunities for international actors that need to measure impact but cannot wait until the end of an intervention to receive much needed information on what is working, what is not, and why.

Evaluation Approaches and their Relation to Impact

Evaluation approaches refer to the principles guiding the design (and implementation) of the evaluation. In other words, they provide 'the framework, philosophy, or style of an evaluation'.[40] There are a large number of evaluation approaches that may frame impact assessment methodologies. This section does not seek to be exhaustive, but rather provides an eclectic overview of some of the main approaches used in the evaluations of international actors as relevant to the assessment of impact in peacebuilding contexts. It should be noted that evaluation approaches are not necessarily mutually exclusive, and can often be used in a complementary manner.

Table 1: Overview of Approaches Examined

Type	Approach	Application
Attribution	Scientific-experimental	Claims attribution through use of counterfactual analysis
Contribution	Theory-based	Supports contribution by testing assumptions at each level of the theory of change[41]
	Participatory	Supports contribution by listening to perceptions of the beneficiaries of what initiatives have made a difference in their lives
Non-causal	Action evaluation	Supports the collective definition of goals – therefore helps to identify jointly what impact should be measured Does not support attribution or contribution
	Goal-free evaluation	Examines the 'actual' impacts of an intervention by deliberately avoiding knowledge of the intended goals and objectives of the project team Does not support attribution or contribution
	Results-based evaluation	Seeks to measure impact to the extent that it focuses on that level of the results chain (i.e. with the use of indicators) Does not support attribution or contribution
	Utilisation-focused	Can address impact depending on methods and the designated use of the evaluation Does not support attribution or contribution

The approaches discussed are separated in Table 1 according to how they relate to measuring impact: whether they can demonstrate attribution, show plausible contribution, or can help to focus the evaluation on impact without necessarily seeking causality.

A range of approaches used by international actors are amenable to measuring impact. The approach claimed to provide the strongest causal link between the intervention and the impact is the *scientific-experimental approach*.[42] It is often considered as the 'gold standard' in assessing impact due to its ability to show attribution. It is based on counterfactual analysis that enquires what would have happened to the welfare of the

beneficiaries if the intervention had not taken place. Based on statistical methods, comparison groups are established to examine the difference between those who received support from an intervention and those who did not (e.g. randomised control trials). This difference is then considered as impact, based on statistically rigorous quantitative measurement techniques. While this approach is considered standard in the development field, its relevance in post-conflict settings is increasingly questioned.[43] This is because complex settings often include a multitude of actors and a number of influencing factors, which make attribution difficult. Moreover logistical challenges such as the difficulty of identifying control groups in multi-site interventions and associated costs are another concern. Finally, the intention to attribute impact rather than only recognize contribution raises ethical questions in so far as such efforts may hamper national ownership. There are also ethical concerns in justifying the limitation of the intervention to a specified group when the control group has been affected by the same factors that define the group receiving the assistance.[44]

Given the challenges associated with this approach, techniques that seek to demonstrate contribution as opposed to attribution have gained increasing attention. These approaches aim to show the contribution of an intervention but recognise that there are many other factors and actors at play contributing to its positive or negative impact. These include participatory methods and theory-based approaches.

Theory-based approaches focus on the underpinning assumptions (theories of change) implicit or explicit in a programme design, and aim to test these. These approaches thus seek to identify 'why' and 'how' changes take place in a project, by identifying, articulating and testing the theory of change that links the results chain of a project (from outputs to outcomes and impacts).[45] Such evaluations go beyond conventional results-based approaches by testing whether an intervention was unsuccessful due to implementation flaws or inconsistent theories.[46] Theory-based evaluations are useful for both accountability and learning purposes as they provide valuable conclusions on what works, what does not and why.

Participatory approaches seek to involve the intervention's primary beneficiaries directly in the planning, monitoring and evaluation process.[47] This may be based on the use of participatory methods such as interviews, focus group discussions and workshops. Advocates of this approach to measuring impact consider local stakeholders to provide 'the single most

valuable source of insights' on the impact of an intervention.[48] By listening to the perceptions of the beneficiaries on what initiatives have made changes in their lives, it is the local communities that can provide evidence of cause and effect.[49] Participatory approaches are also recognised as extremely valuable in terms of capacity building, promoting local ownership and enhancing the use and relevance of the evaluation itself. While there may be concerns over the possibility of biased findings due to conflicts of interest among local/national stakeholders,[50] different levels of participation can be supported depending on what methodology is used.

Finally, there are certain other non-causal approaches to evaluation, which although they cannot demonstrate attribution or contribution may still support evaluations by making them more 'impact-focused'. These approaches include action evaluation, goal-free evaluation, results-based evaluation and utilisation-focused evaluation.

Action evaluation is based on the evaluator facilitating, together with the project team and key stakeholders, the identification of goals and objectives throughout the life of a programme. Progress towards the achievement of the goals/objectives is monitored and later evaluated jointly by the evaluator and project team. Essentially, this approach supports the collective identification of context-specific criteria of success according to which an intervention is then evaluated.[51] The advantages of this approach are that it promotes joint understanding of success, enables adaptation to changing environments and can be used to track impact. However, this may require a significant change in mind-set if an action-oriented approach seeks to allow deviation from its original goals. Also, enabling local stakeholders to define the evaluation criteria may hamper the ability to evaluate an intervention against a mandate set at the international level. The value of this approach lies in its ability to define a joint vision of impact (goals) rather than its role in data collection and analysis.[52]

Goal-free evaluation approaches aim to examine the *actual* outcomes and impacts of interventions in contrast to those intended. The external and independent evaluator/team thus deliberately avoids prior knowledge of the intended goals and objectives of the intervention and has only minimal contact with the project team. The focus is on the actual results of the intervention rather than on the intended results.[53] The aim is to uncover the positive and negative effects of an intervention – whether

intended or not – by limiting the bias of the evaluator. The evaluator collects information about programme results and then compares these with the actual needs of the beneficiaries in order to determine effectiveness.[54] The data collection process of a goal-free evaluation can be time-intensive and more expensive than conventional results-based evaluations, since the evaluator/team has to investigate a very broad range of issues, often using participatory methods involving the primary beneficiaries and other key stakeholders.[55] A great advantage of this approach is that it can demonstrate impact without needing a coherent logical framework that sets out the relationships between inputs, outputs, outcomes and impact. It is therefore of particular use where the programme logic is weak, as may often be the case in fast-paced conflict environments.[56]

Results-based evaluation is the approach most widely adopted by the international community. This approach seeks to identify to what extent the intended broad goals and specific objectives of an intervention have been met according to defined indicators, benchmarks and baseline studies. It can focus on impact only to the extent that indicators at this level are developed. While this approach does not support contribution of an intervention to an impact, it can highlight the linear relationships between outputs, outcomes and impact. Results-based approaches do not seek to evaluate the utility of the objectives or goals set for the intervention.[57]

Utilisation-focused evaluation is based on the understanding that evaluations 'should be judged by their utility and actual use'.[58] This involves a process whereby the evaluator helps users select the most appropriate approach and methods for the purpose of their evaluation.[59] These evaluations can thus use any type of design and methods, and can focus on any level of the results chain.[60] Such evaluation is therefore not specifically associated with impact, but could be adapted to measuring impact if a project team were to jointly decide on the most appropriate methods for the context at hand. The underpinning idea is that the utilisation-focused evaluation approach deepens the feeling of ownership of the project team in the evaluation process through their active involvement, and increases the utility of the evaluation – i.e. it becomes more likely that the team will actually use the evaluation results and these will effect change in the project.[61] This could be of importance in the area of measuring impact where there may be reluctance to adapt to new evaluation approaches.

This approach has often been advocated by UN entities, possibly when faced with resistance to evaluations that are perceived as externally imposed.

Methodologies for Measuring Impact

These broad approaches to evaluation are applied in practice through a number of methodologies. Evaluation methodology refers to 'the term covering the different methods to be applied to meet the overall purpose and objectives of the evaluation. The particular methodology to be used for data collection and analysis is determined by the subject and purpose of the evaluation.'[62] Some methodologies are more applicable as a learning tool because they seek to answer questions related to why the intervention had an impact or not while others may be more relevant in order to provide for accountability. The appropriate evaluation methodology should therefore be determined in view of its ability to answer the questions that the evaluation is seeking to answer.

Six methodologies seem applicable to measuring impact with particular relevance for rule of law and security institutions: a) impact evaluation; b) theory-based impact evaluation; c) contribution analysis; d) outcome mapping; e) RAPID outcome assessment, and f) most significant change (MSC). All of these methodologies have in common that they can be used to illustrate impact. However, they are all different in terms of the evaluation questions they can most usefully answer, and in terms of the time, cost and skills-sets required to use them. Moreover, some of these can be used independently while others must be combined with other methodologies in order to provide a fuller picture for evaluation purposes. The following table provides an overview:

Table 2:   Overview of Methodologies for Measuring Impact

	Methods	**Relevance**
Impact evaluation (IE)	Quantitative methods such as control groups (e.g. randomised control trials) and before/after comparisons	Enables attribution by undertaking a 'counterfactual' analysis to compare what actually happened with what would have happened in the absence of the intervention - quantifies impact
Theory-based impact evaluation (TBIE)	Quantitative and qualitative methods - Control groups and before/after comparisons combined with theory of change approaches	Strengthens traditional impact evaluation by using theory of change to understand what worked, what did not, and why
Contribution analysis	Qualitative methods such as case studies, MSC stories, focus group discussions	Seeks to show plausible evidence of effect of an intervention by testing programme logic and theory of change
Outcome mapping (OM)	Qualitative methods such as focus group discussion, workshops and use of 'progress markers'	Focuses on measuring change based on the premise that changes in behaviour of key stakeholders will ultimately contribute to impact
RAPID outcome assessment (ROA)	Draws on outcome mapping methodology, MSC technique and episode studies to enable triangulation of data	Seeks to assess and map the contribution of a project's actions to a particular change in policy or the policy environment
Most significant change (MSC)	Qualitative methods such as group discussions, interviews and workshops to support systematic selection of MSC story	Collection of 'significant change stories' that are perceived as being the most significant in contributing to impact on people's lives: Illustrates change rather than measuring impact *per se*

Impact evaluation

Impact evaluations aim to measure and establish the value of impacts that can be *attributed* to an intervention. The methodology is based on the

definition of cause-effect hypotheses to be tested and the use of quantitative methods for data collection and analysis.[63] This is done through a 'counterfactual' analysis of the impacts of an intervention, i.e. a comparison of what actually happened with what would have happened in the absence of the intervention. Impact evaluations employ experimental and quasi-experimental quantitative methods. They are useful for answering evaluation questions relating to 'whether development interventions do or do not work, whether they make a difference, and how cost-effective they are.'[64]

Impact evaluations are usually *ex post*, meaning that they are conducted some time after the end of an intervention. There are two main components:

- Control groups are randomly selected along with treatment groups before the intervention starts. As a result, the differences between the two groups in terms of impacts can be attributed to the intervention without the risk of selection bias.[65] The most popular variant of this method is *randomised control trials* (RTC).[66] Quasi-experimental techniques are applicable in cases where the groups were not randomly assigned.

- Before/after comparisons assess the situation at the time of the evaluation with the situation before the intervention. In complex environments, especially post-conflict settings, baseline data for comparisons are very frequently lacking, making this approach less applicable.

Applying the Impact Evaluation methodology generally involves three key phases:[67]

1. *Preparation:* Formulation of cause-effect hypotheses for testing and identification of the control group.

2. *Implementation:* Data collection (e.g. questionnaires, standardized interviews).

3. *Analysis:* Statistical analysis of the data followed by interpretation of the results.

An impact evaluation is generally very time intensive and the cost can be 'significant'.[68] The actual collection of data and observation of impacts can take years, depending on the intervention and the effects to be observed.[69] The statistical data analysis may take several months. Finally, impact evaluations require very sophisticated scientific skill-sets. A drawback of impact evaluation is that it cannot highlight why events evolved the way they did because it focuses more on quantifying impact. In order to adapt to the growing demand to understand how impacts emerge, theory-based impact evaluation has been developed as a method that combines theory of change methods with quantitative techniques (see below).[70]

Theory-based impact evaluation

Theory-based impact evaluation (TBIE) takes traditional impact evaluation methods a step further by combining them with theory of change methods to generate better understanding of the underlying reasons for success (or failure). Applying the TBIE approach increases the policy relevance of scientific-experimental impact evaluations [71] by going beyond the determination of an impact to interpret and evaluate the findings.[72] Such evaluations aim not only to establish what impact a program has had but also to "understand *why* a program has, or has not, had an impact".[73]

   Mixed methods (qualitative and quantitative) for data collection and analysis are used.[74] Building on the quantitative counter-factual based techniques described above, complementary qualitative methods may include for example: Document analysis (for example, analysing project documents to understand the project logic; academic and political literature to understand the context and inform the evaluation design), focus group discussions, various types of interviews with project staff, beneficiaries and key stakeholders.

The application of the TBIE approach is based on six steps:[75]

1. *Mapping the causal chain (programme theory):* This means mapping the connection between inputs, outcomes and impacts, in order to reveal the theory of change underlying the programme logic.

2. *Context analysis:* This step involves conducting a desk study of project documents and relevant academic literature prior to

designing the evaluation. It should support an awareness of the social, political and economic factors that may affect the context of the program or its evaluation.

3. *Anticipating heterogeneity of impact:* This involves analysing factors that may affect impact such as the socio-economic context, the kind of beneficiaries targeted and the design of the intervention. This is important for identifying both the appropriate sample size and the sub-groups to be tested before an evaluation takes place.

4. *Establishing a credible counterfactual:* A counterfactual should be identified using a control group carefully selected to avoid selection bias. Considerations include the risk of the control group being affected by the intervention in question ("spill over effects") or by other interventions ("contamination" or "contagion").

5. *Rigorous factual analysis:* Factual analysis to understand the logic behind the causal chain and potential breaks in the chain that may have resulted in low impact. This would include target analysis (of the beneficiaries) in order to check whether there were targeting errors.

6. *Use of mixed methods*: This involves considering the balance of qualitative and quantitative approaches to be used.

While this approach is widely accepted in theory,[76] its actual use in impact evaluations has been limited. This may be due to the significant costs in time and resources associated with implementing the methodology, given the need to combine both rigorous scientific methods and detailed analysis of theory of change. The advantage of the methodology is that if scientific methods are warranted, this version enables using the evaluation as a learning tool and not just as a means of accountability. The use of impact evaluation is also limited by the fact that sufficient time must pass before statistically relevant information can be generated.

Contribution analysis

Contribution analysis aims to provide plausible evidence of the difference a programme is making (contributing) to observed outcomes. It does so by assessing the underlying theory of change and disaggregating analysis at each level of the results chain. A performance story is then built about the

contribution the intervention has made based on the analysis of the programme logic, results achieved and having considered alternative causal explanations for the same results (other influencing factors). Contribution analysis is different to traditional results-based tools in that in addition to assessing evidence linking a programme to results, consideration is also given to assessing the assumptions of the theory of change and the influence of external factors and actors. A performance story is developed to explain why it is reasonable to assume that the actions of the programme have contributed to the observed outcomes. Additional evidence is then sought to support the claims. A contribution analysis is usually conducted by an independent, often external, evaluation expert or team. However, most of the aforementioned steps can also be conducted in a participatory way with some involvement of the key stakeholders and beneficiaries.[77] For example, the performance story can be submitted to relevant stakeholders to test their agreement with it and help identify where the weaknesses lie. A prerequisite for contribution analysis is that the programme or project has been planned and implemented within the logical framework approach, based on a specific theory of change.

The methodology employs a mixed-methods approach for data collection and analysis. In addition to quantitative and formal data sets, it is suggested that the evaluator collect qualitative data including for example case studies based on documentary analysis, various types of participant interviews or use of anecdotal and informal stories (as used in Most Significant Change technique).[78]

Contribution analysis involves six iterative steps:[79]

1. *Acknowledge the attribution problem.* This includes recognizing that there are problems related to attribution; determining the specific cause-effect question to be addressed; exploring the type of contribution that can be expected; identifying the other factors and actors that could influence the outcomes of the intervention; and assessing the plausibility of the expected contribution.

2. *Build a theory of change.* This includes developing a programme logic that details how the intervention was intended to work; listing the assumptions underlying the theory of change; and assessing whether and how the theory of change can be challenged.

3.   *Gather evidence on the theory of change.* This includes seeking evidence that the intervention was implemented as originally planned and determining to what extent the assumptions in the theory of change are valid.

4.   *Assemble and assess the contribution story.* Develop a performance story of why it is reasonable to assume that the actions of the program have contributed to the observed outcomes. This includes determining the strengths and weaknesses of the story and assessing to what extent stakeholders agree with the story.

5.   *Gather additional evidence.* Gather additional evidence to improve the program's performance story. This can involve seeking information on both the extent of occurrence of specific results in the results chain and the strength of certain links in the chain.

6.   *Revise and strengthen the contribution story.* Revision and strengthening of the performance story in order to enhance its credibility.

This methodology is easy to use, building on the logical framework approach. It requires additional information in the logical framework from the outset to support the subsequent testing of a clear theory of change for each level, the assumptions of the theory of change and potential influencing factors.[80]

Outcome mapping

The originality of outcome mapping is that instead of aiming to measure impact, it focuses on outcomes in terms of behavioural change of individuals, groups and institutions and their relationships.[81] Thus outcomes are defined as 'changes in the behaviour, relationships, activities, or actions of the people, groups, and organizations with whom a program works directly'.[82] Outcome mapping does not seek to measure impact in terms of 'tangible products', but is based on the understanding that changes in the behaviour of key stakeholders and partners will ultimately contribute to impact.[83]

Context-specific progress indicators (so-called *progress markers*) are used to document the progress of the intervention towards the desired

outcomes. They can be seen as 'a graduated ladder of specific changes in boundary partner behaviour and relationships which define and describe progress towards each outcome challenge.'[84] The value of these markers is that they can capture types of change that are hard to measure with more 'traditional' indicators that need to meet criteria referred to as SMART: specific, measurable, achievable, relevant and time-bound. The data collected for monitoring and evaluation purposes are mostly qualitative and anecdotal in nature. To trace changes and progress, outcome mapping suggests the use of journals (e.g. to monitor relationships) together with other data collections methods such as: different types of interviews including group techniques (interviews, focus group discussions, workshops), and document reviews. [85]

The outcome mapping process contains three stages: [86]

1. *Planning phase*: The aim is to clarify and reach consensus on what behavioural changes the project or program desires to achieve. It seeks to design the intervention to achieve these outcomes and decide jointly on how to measure these outcomes (including agreement on a strategy and performance indicators). This is usually done in one or several workshops facilitated by an external expert.

2. *Outcome and Performance Monitoring.* This stage involves monitoring the project's progress and the stakeholders' and partners' progress toward the achievement of defined outcomes by reporting on changes in the actions of the actors involved. The M&E framework for data collection and reporting is based on the use of journals: an outcome journal (monitors actions and relations among boundary partners), a strategy journal (monitors activities and strategies) and a performance journal (monitors organisational practices).

3. *Evaluation Planning.* When outcome mapping is used primarily as a monitoring tool, this stage serves to identify evaluation priorities and come up with an evaluation plan.

In theory outcome mapping relies on data generated by the programme team and boundary partners, which means that the team must have the time to dedicate to such activities, and the boundary partners (national stakeholders directly linked to the programme) must also be

willing to engage actively with the evaluation. In practice however, rather than requiring boundary partners to collect the information themselves, the approach can be adapted for use with exclusively external assessments, for example by using questionnaires that examine progress markers.[87]

The OECD DAC handbook on security sector reform lists outcome mapping as a potential method for measuring impact in cases where 'changes in attitudes and behaviours (e.g. among senior officials working in security and justice institutions) [are] just as important as any specific practical changes that occur'.[88] While outcome mapping alone is unlikely to be sufficient for conducting evaluations at the impact level, elements of this methodology can be used specifically to analyse behavioural change and then combined with other tools as part of a broader evaluation approach: For example, progress markers could be developed to supplement regular indicators. Outcome mapping can also be combined easily with the logical framework approach.

RAPID outcome assessment

RAPID outcome assessment (ROA) aims to 'assess and map the contribution of a project's actions on a particular change in policy or the policy environment'.[89] It is a systematic approach to collecting data in order to analyse how changes in the behaviour of key stakeholders contribute to policy changes. With the aim of learning, ROA is a flexible and adaptable methodology that can be used in combination with other M&E approaches.[90]

ROA draws on the outcome mapping methodology, the MSC technique and episode studies. In doing so, ROA enables triangulation of data which can therefore increase the reliability of findings.[91] MSC provides tools to identify and rank key changes, while episode studies provide ROA with a toolset to work backwards from an observed policy change (result) to find the most relevant factors that contributed to it.[92] The information gathered for the assessment is of a qualitative nature. Generally it involves background research, a workshop to identify change processes and the triangulation of data. Methods for data collection and analysis may include (the list is not exhaustive): participatory workshops, key informant interviews, narrative and anecdotal stories on contributions to key policy changes, and document analysis.

The ROA methodology has three main stages:[93]

1. *Background research and preparation:* This stage involves document review and dialogue in order to understand the intervention's history and intended changes.

2. *The ROA workshop:* A workshop is held with stakeholders to identify the key policy change processes. Timelines are developed in order to show how behaviour has changed in line with changes in the project and external factors.

3. *Triangulation and refining conclusions:* This stage should enable the refining of the stories of change on the basis of the material collected during the first two stages. The aim is to identify the key policy actors, events and their contribution to the observed changes. Follow-up interviews to refine conclusions can be conducted.

The RAPID Outcome Assessment can be done internally by the project team or by an independent external expert. It is a participatory methodology, which involves all key stakeholders and project partners in all stages of the process.

Most significant change (MSC)

The MSC technique is a participatory monitoring and evaluation methodology, originally developed to illustrate change processes in community-based development projects. [94] The methodology aims at collecting so-called 'significant change stories', and systematically selecting the one considered to have had the most significant impact on people's lives. The MSC approach uses participants' stories of change to spot the outcomes and impacts (intended and unintended, positive and negative) of an intervention and classify them in a hierarchical way through ongoing dialogues between the project team and the beneficiaries. It is important to highlight that MSC does not aim to measure impact with predetermined indicators, but rather to illustrate change.[95] The data collection, monitoring and analysis are done by those most directly involved – participants and field staff.

The main methods for data collection and analysis may include common tools of participatory M&E approaches, such as: focus group discussions, different types of interviews and workshops. The implementation of the MSC technique generally involves the following steps, although they can be adapted to fit the needs of a specific project:[96]

1.   *Introduction:* This stage involves raising interest, defining the domains of change, and defining the reporting period.

2.   *Implementation:* This stage involves the collection of significant change stories, followed by the selection of the most significant of the stories.

3.   *Analysis:* The stories are verified and quantified. Secondary analysis and meta-monitoring is undertaken before the system is repeated and revised.

MSC can be useful in challenging contexts where a 'real-time' impact assessment is needed to monitor the immediate effects the intervention is having on the lives of beneficiaries.[97] It is also considered useful in situations where there is more emphasis on significant change being achieved than on specific original objectives being met.[98] However, it requires 'an organizational culture where it is acceptable to discuss things that go wrong as well as success, and a willingness to try something different.'[99] If applied as intended, that is, as an iterative and participatory method, it is a time-consuming approach. But it can also be simplified and adapted for use by external evaluators to supplement other evaluation methods by including questions on 'most significant change' in questionnaires or focus group meetings.

Summary

This section has demonstrated that there is a wide range of methodological approaches that can be used for measuring impact, ranging from scientific-experimental designs to participatory or utilisation-focused designs. There is also a range of impact assessment methodologies that can be used to

Table 3: Methodologies and Approaches [100]

Methodologies / Approaches	Impact evaluation	Theory-based impact evaluation	Contribution analysis	Outcome mapping	ROA	MSC
Scientific-experimental	X	X				
Theory-based		X	X	X	X	
Participatory		/	X	X	X	X
Action-oriented				X	X	X
Results-based			/			
Utilisation-focused	/	/	/	/	/	/

measure impact. Table 3 below shows how the impact assessment methodologies fit within the approaches discussed. This is intended to highlight which methodologies correspond to approaches that international actors would like to incorporate in how they measure impact.

The Goal-free evaluation method is the only approach where no corresponding methodology for measuring impact was found within the limits of this study, and it is therefore not reflected in the table. However, it can be considered both an approach and a methodology since the principle guiding the design of the evaluation is that there is no knowledge of the originally intended goals. This is also how the evaluation is intended to be undertaken in practice – the methods used as part of the methodology are therefore based on the need to collect information from the beneficiaries while avoiding prior knowledge of the goals. The utilisation-focused approach can be applicable to any methodology, as it depends on the purpose of the evaluation and the decision of the project team.

Impact evaluation is a scientific-experimental approach to evaluation. Since Impact Evaluations are usually conducted ex-post, several years after the completion of a project, they are generally not conducted in a participatory manner.

Theory-based impact evaluation combines scientific-experimental designs with a theory of change approach. It also draws on qualitative data,

provided to a certain extent by participatory methods, to enable the analysis of the underlying project assumptions and theories of change.

Contribution analysis is a participatory approach to the extent that the external independent evaluation team gathers the perceptions of key stakeholders and beneficiaries to test the theory of change; but it does not belong to the type of participatory approach that enables national capacity building through evaluation. Its participatory component can be strengthened by submitting a performance story to relevant stakeholders to test their agreement with it and potentially help identify where weaknesses lie. It is also a theory-based approach in so far as the concept of theory of change is a crucial component. It builds on the results-based approach, but it is not a results-based approach *per se* as it requires further information on the underlying assumptions of the theory of change and assessment of the influence of external factors.

Outcome mapping is a participatory approach that involves all key stakeholders and to a certain extent the beneficiaries in the planning, monitoring and evaluation process. Similar to MSC methods, outcome mapping processes are implemented within an action-orientated approach: the project team and key stakeholders are assisted by an external facilitator in setting objectives and goals for behaviour change in the planning phase of the project, and decide on ways to measure and evaluate the jointly agreed goals. Also, MSC and outcome mapping feature another characteristic of action-orientated approaches, namely the possibility to adapt the project and the corresponding monitoring and evaluation system to changing circumstances over the course of implementation. Finally outcome mapping can be considered as a theory-based approach, since it aims to understand the change processes a project has contributed to.

RAPID outcome assessment is a combination of both MSC and outcome mapping and therefore, it is also considered to be a participatory as well a theory-based and an action-orientated approach.

Most significant change is a genuine participatory approach in the sense that it relies on the active engagement of stakeholders in the evaluation as well as in the planning and monitoring phase. It can also be considered as an action-oriented approach to the extent that stakeholders (especially the project team) come together to define their criteria of success by contributing stories of what they have considered to be most important elements of impact.

This section has highlighted that each methodological approach and impact assessment methodology can serve different purposes. Some approaches are more suited to supporting the participation of national actors (e.g. participatory and action-orientated evaluation approaches) while others are more fitted for testing the theory of change (e.g. theory-based approaches). Similarly, some methodologies are better suited for measuring behaviour change (e.g. outcome mapping) while others are better geared to answer questions related to the extent of impact achieved (e.g. impact evaluation). They all have different strengths and weaknesses. Moreover, they are not mutually exclusive. Most of them can be combined and used in a complementary manner. Most can also build on traditional results-based management approaches used by many international actors. Building on this discussion the next section turns to examine which of these approaches and methodologies international actors are using in practice to measure their impact.

MEASURING IMPACT: APPROACHES OF INTERNATIONAL ACTORS

This section examines the current approaches taken by international actors to evaluations and how these relate to measuring impact.[101] It looks at the challenges actors face in addressing the need to measure impact and how they seek to overcome them. The discussion considers both bilateral and multilateral actors, with emphasis on UN actors. The actors covered by the analysis are those currently most proactive in the development of approaches and methods for evaluation. These actors have also made substantial information on their approaches to measuring impact publicly available.

The discussion considers the following bilateral actors:

- Australian Agency for International Development (AusAID)
- Canadian International Development Agency (CIDA)
- United Kingdom Department for International Development (DFID)
- Japan International Cooperation Agency (JICA)
- Swedish International Development Cooperation Agency (Sida)
- United States Agency for International Development (USAID).

Multilateral actors discussed include:

- African Development Bank (AfDB)
- European Commission Development and Cooperation (EuropeAid)

- Organisation for Economic Co-operation and Development (OECD)
- World Bank
- United Nations.

UN actors are examined separately because it is important to highlight the differences in approaches to measuring impact within the UN system itself. The UN entities examined are members of either the UN Rule of Law Coordination and Resource Group or the UN Inter-Agency SSR Task Force.[102]

UN entities examined include:

- Department of Peacekeeping Operations (DPKO)
- Office of the UN High Commissioner for Human Rights (OHCHR)
- Peacebuilding Support Office (PBSO)
- United Nations Children's Fund (UNICEF)
- United Nations Development Programme (UNDP)
- United Nations High Commissioner for Refugees (UNHCR)
- United Nations Entity for Gender Equality and the Empowerment of Women (UN Women)
- United Nations Office on Drugs and Crime (UNODC).

The evaluation approaches of international actors examined are mostly generic – that is to say, they are broad approaches outlined in policies and guidelines that are applicable to a number of activities the actors are engaged in including but not limited to peacebuilding and support to rule of law and security institutions. The evaluations and meta-evaluations examined in this section can be grouped into two main categories: cross-cutting generic (e.g. country programmes) and sectoral (e.g. health, gender, development). While the information from this section is therefore not specifically linked to rule of law and security institutions issues (due to the scarcity of available impact assessments focusing on these issues), there are several insights that can be drawn which are nonetheless relevant to measuring the impact of peacebuilding support for rule of law and security institutions.

Bilateral Actors

All the bilateral actors examined have recognised the importance of measuring impact in their internal policies and/or guidelines. In some cases this is for reasons of accountability: for example, AusAID is required to demonstrate impact in informing the government's budget process.[103] In other cases it is due to the recognised learning potential: for example, according to JICA's annual evaluation report, impact evaluation is considered to be important because simpler methods that compare outcomes before and after project implementation have a tendency either to over- or underestimate change.[104]

Many of the bilateral actors' policies and guidelines consider approaches that enable attribution as being the most 'rigorous'. For example, USAID considers the best impact evaluations to be those 'in which comparisons are made between beneficiaries that are randomly assigned to either a treatment or a control group'.[105] USAID requires all projects using untested hypotheses or new approaches to undergo impact evaluations where possible; if this is not possible, a detailed statement explaining why must be included in the final report and a performance evaluation can be done instead.[106] JICA also promotes methods that enable attribution, specifically undertaking a trial of four different techniques within impact evaluations.[107] These were all tested in an attempt to find alternatives to randomised control trials, which were noted as being especially difficult to conduct in certain fields.[108]

The extent to which advocates of experimental approaches actually use them in practice is questionable. Evaluation reviews often point to measuring impact as a weak component of evaluation approaches. For example, while USAID advocates experimental methods to evaluate impact, a sample of evaluations examined revealed that while a few took a statistical/quasi-experimental approach, the majority did not. Moreover, a meta-evaluation of its recent evaluations noted that only 26 per cent of evaluations used the logic model in order to determine the causal relationship between inputs, outcomes and impacts.[109] A similar finding is reflected in a Sida assessment of 34 evaluation reports in 2008, which found that none used experimental or quasi-experimental methods.[110] Most of the data came from document reviews and open-ended

interviews.[111] The Sida study highlighted that less than half the reports provided a satisfactory analysis of impact.[112]

The Independent Commission for Aid Impact (ICAI), which is an important player in evaluating UK aid impact, notes that in general statistical methods such as randomised control trials are not to be used due to the time they take to implement and the fact that they would need to be built into the project design.[113] However, where there are concerns about the results of a project, quantitative analysis may be used, including statistical techniques that 'construct a suitable "counterfactual" in place of a control group, so as to test for attribution more rigorously'.[114] In practice, several reviews of DFID evaluations concluded that M&E were 'not sufficiently focused on impact' and 'not able to attribute results to DFID support'.[115] Challenges in establishing counterfactuals have been noted, particularly concerning the need to identify a counterfactual from the outset of the intervention as opposed to searching for one retrospectively.[116]

The majority of actors therefore promote experimental and quasi-experimental approaches but also note that these may be too challenging for their needs. Challenges raised relate to logistical obstacles: for example, proving attribution has been raised by CIDA as a challenge due to the difficulty in finding a robust methodology that is feasible to use.[117] Sida's Evaluation Manual acknowledges that a plausible counterfactual may be difficult to establish, and in those situations weaker arguments based on expert knowledge would be acceptable.[118] Similarly an AusAID evaluation also pointed to the difficulty of relying on impact evaluation due to the fact a lack of baseline and parallel activities rendered attribution difficult.[119] Cost-related disincentives have also been raised: for instance, AusAID's best practice brief recognises that the greatest cost-benefit in the area of impact assessment can be achieved by limiting itself to a 'handful' of experimental or quasi-experimental impact evaluations.[120]

There is a general recognition among bilateral actors that in practice it may be more realistic to rely on 'best extent possible' approaches that seek to show contribution by other methods.[121] Such methodologies increasingly being used by international actors include outcome mapping, contribution analysis and MSC while few examples were found of ROA. Concerning the use of qualitative methods more generally, AusAID in

Table 4:   Examples of Bilateral Actors' Approaches to Measuring Impact

	Approaches advocated in theory	Approaches demonstrated in practice[122]
AusAID	• Impact evaluations using both qualitative and quantitative methods, randomly chosen subjects, counterfactuals and statistically simulated counterfactuals[123]	• Recent evaluation reports all address impact in a limited way;[124] in practice, impact was sometimes measured anecdotally[125] and sometimes deemed impossible[126] • Impact evaluation of a HIV/AIDS project.[127] • Contribution analysis was used in Fiji[128] and Papua New Guinea in the law and justice sector[129] • MSC methodologies used in Papua New Guinea[130]
CIDA	• Commitment to the International Initiative for Impact Evaluation • Attribution is considered a challenge, so a 'best extent possible' approach is advocated	• Multi-country impact evaluation conducted for a large telecommunications project[131]
DFID	• Randomised control trials are advocated where possible and ethical[132] • Qualitative approaches, including participatory ones, also advocated for measuring impact[133]	• Impact evaluations relating to health projects in Malawi[134] and Bangladesh[135] • Outcome mapping used, e.g. for climate change adaptation projects in Africa[136] • RAPID outcome assessment used at least once to measure impact on health sector[137]
JICA	• Participatory evaluation is promoted, based on stakeholder-based evaluation, utilisation-focused evaluation and empowerment evaluation[138]	• Impact is measured six months prior to completion of project or *ex post* for up to seven years following completion • Currently devising methods for measuring impact • Impact evaluation piloted with four irrigation projects[139]

Sida	• Advocates scientific-experimental measurement of impact, but recognises that other methods may be more realistic • Promotes utilisation-focused and participatory approaches	• Impact assessment conducted in Belarus[140] • Outcome mapping[141]
USAID	• Observational, quasi-experimental and experimental methods should all be considered[142] • Requires all projects using untested hypotheses or new approaches to undergo impact evaluations; advocates methods which use control groups	• Impact evaluations conducted in Africa[143] and Serbia[144]

particular recognises their value specifically in addressing the reality that AusAID programmes are often delivered over short timeframes. To meet this challenge alternative approaches were sought that would show progress towards results as opposed to establishing rigorous causal links. On this basis for example contribution analysis was selected as a methodology to evaluate the impact of an education programme in Fiji.[145] Australia is among the first bilateral donors to use contribution analysis in its evaluations.[146] Among others, it was found to have supported donor harmonisation by focussing on developing alternative explanations for progress and thereby allowing a greater appreciation of other donor activities and supporting synergies.[147] Moreover, contribution analysis was also considered to generate significant improvements in programme logic.[148] Sida has also used outcome mapping in an evaluation of civil society projects in Bosnia-Herzegovina, and recognised its benefits over the logical framework approach.[149]

Multilateral Actors

The multilateral actors examined have all recognised the importance of measuring impact. The AfDB, the OECD's Development Assistance Committee (OECD DAC) and the World Bank are all members of NONIE, a 'Network of Networks for Impact Evaluation comprised of the OECD DAC

Evaluation Network, the United Nations Evaluation Group (UNEG), the Evaluation Cooperation Group (ECG), and the International Organization for Cooperation in Evaluation (IOCE) – a network drawn from the regional evaluation associations'.[150] A key element is the focus on experimental design as the main approach to measuring impact. NONIE is one of the lead actors promoting the 'impact evaluation' methodology outlined in the previous section, and is a proponent of attribution analysis based on the counterfactual. It does note that in certain cases contribution analysis may be the most realistic approach, but it only addresses alternative methods in its annex. In its annex, there is one page on qualitative methods for assessing the effects of interventions including outcome mapping and MSC among others.[151] Outcome mapping and MSC are methodologies that, as discussed above, are relevant to international actors engaged in support to rule of law and security institutions. The following provides examples of the approaches relevant multilateral actors have taken in practice:

Table 5: Examples of Multilateral Actors' Approaches to Measuring Impact

	Approaches advocated in theory	Approaches demonstrated in practice
AfDB	• Focus on impact evaluations is emerging[152] • Promotes attribution with 'before-and-after' methods • Member of NONIE[153]	• Several impact assessments carried out in areas of poverty alleviation,[154] energy sector[155] and water supply[156] although they do not all use quantitative methods • MSC[157]
EuropeAid	• Guidelines mention impact[158] • Counterfactuals, contribution analysis and attribution analysis are mentioned[159]	• There appear to be no reports or sections of reports addressing impact in isolation • Isolated examples of most significant change[160] and contribution analysis[161]. • Impact is mentioned as a criteria in many evaluations although they are not explicitly impact evaluations[162]

OECD	• Impact evaluation[163] • Member of NONIE	• Evaluations conducted by the aid agencies of member states[164]
World Bank	• Overall evaluation approach is objective-based[165] but also undertakes scientific-experimental evaluations • Member of NONIE with a similar theoretical approach as OECD	• Multiple impact evaluations available online[166] including quasi-experimental[167] and non-experimental[168]

The AfDB promotes self-evaluations that assess attribution of socio-economic and policy changes.[169] The guidelines suggest that modified 'before-and-after' methods should be used to deduce attribution based on both quantitative and qualitative data.[170] In practice, it has considered conducting impact evaluations to be challenging at times due to the need for pre-evaluation studies and the lack of available baseline data.[171] The MSC approach has been used for the evaluation of an AfDB decentralisation strategy to identify 'difficult to quantify changes' that are not easily captured by traditional M&E. In this context stakeholder stories were developed to support the testing of the theory of change.[172]

EuropeAid's evaluation guidelines recognise the need to focus on attribution and contribution. The guidelines put forward four strategies: change analysis, meta-analysis, attribution (counterfactual) analysis and contribution analysis.[173] While the choice of analysis is left to those conducting the evaluation, the guidelines mention the usefulness of the latter three strategies when answering cause-and-effect questions such as measuring the EC's attribution or contribution to an effect, the sustainability of that contribution and whether it is being achieved at a reasonable cost.[174] It has used both MSC and contribution analysis in its evaluations.

The OECD DAC has a subsidiary body called the Network on Development Evaluation. The network's overall goal is to 'increase the effectiveness of development-cooperation policies and programmes by promoting high-quality, independent evaluation'.[175] The OECD DAC definition of impact developed by the network is the most widely used and accepted in the field of international development cooperation, as well as in peacebuilding.[176] In the area of development the network promotes

rigorous impact evaluation, as a member of NONIE.[177] In the area of peacebuilding its Guidance on Evaluating Conflict Prevention and Peacebuilding Activities proposes a further range of different approaches, which can be adapted in peacebuilding settings.[178] Mixed-method approaches are promoted to deal with the complexity of peacebuilding interventions.[179] The guidelines also recognise that in peacebuilding contexts impacts can be longer term or relatively immediate.[180]

The World Bank's independent evaluation group (IEG) works towards quality improvement of impact evaluation.[181] IEG conducts *ex post* impact evaluations several years after the end of a project, using scientific-experimental and quasi-experimental designs.[182] IEG considers that 'good evaluations are almost invariably mixed method evaluations', and therefore uses qualitative data to inform the design and interpretation of quantitative impact evaluations.[183] Both experimental and non-experimental designs have been used in practice to create counterfactual groups (see Table 4).

UN Actors

Evaluation within the UN system takes place at different levels and follows a diverse set of arrangements. There are some UN entities that have specific evaluation roles, such as the Joint Inspection Unit, the Office of Internal Oversight Services and the UN Evaluation Group. While the JIU has no mandate for measuring impact, the OIOS can measure the impact of Secretariat programmes.[184] For OIOS, impact refers to 'the ultimate, highest level, or end outcome that is desired.'[185] It advocates three types of evaluation design, namely experimental (using counter-factuals), quasi-experimental and non-experimental.[186] According to its Inspection and Evaluation Manual, '[n]on-experimental designs are the most commonly used evaluation design' by its evaluation division.[187] In general, it has been recognised that measuring impact is a major weakness in evaluations across the UN system, which can account for the increasing efforts to focus on this issue.[188] The need to measure impact has increasingly been addressed in UN reports. A 2009 report of the Committee for Programme and Coordination recommended that the General Assembly request the Secretary-General to ensure that relevance and impact are the major focus of self-evaluations.[189] Moreover, it was recommended that the various secretariats are made aware of 'the imperative of impact assessment'.[190]

Table 6 outlines what methods UN entities are increasingly promoting for the measurement of impact, and how they have addressed this issue in practice.

Table 6: Examples of UN Entities' Approaches to Measuring Impact

	Approaches advocated in theory	Approaches demonstrated in practice
DPKO	• N/A	• Impact assessment to measure the ten-year impact of UNSCR 1325 in peacekeeping based on a participatory approach [191]
OHCHR	• Results-based management with a focus on evidence-based reporting and contribution analysis[192]	• N/A
PBSO	• M&E approach under development • Approach will be based on understanding that results should be 'attributable'[193]	• N/A
UNDP	• Impact evaluations mentioned in M&E handbook although challenges are recognised[194] • Contribution and attribution mentioned but not methodology[195]	• Joint or UNDP country office-commissioned impact assessments[196]
UNFPA	• Participatory evaluations are advocated	• Recognised as weak in UNFPA meta-evaluation[197]
UNHCR	• Promotes participatory and utilisation-focused approaches, and is striving to foster a results-based management culture[198]	• Independent impact evaluation commissioned in southern Sudan[199]
UNICEF	• Contribution analysis and participatory performance story reporting are all promoted[200] as are utilisation-focused approaches[201]	• Impact evaluations carried out, e.g. in Mozambique[203] • MSC approaches used in India[204] and the Solomon Islands[205]

	• MSC techniques are also promoted[202]	
UNODC	• Advocates utility-focused and participatory approach	• Considered weak in UNODC report, but efforts being made • MSC used for a criminal justice reform programme[206] and an anti-drug-trafficking coordination centre[207]
UN Women	• Participatory approaches advocated[208] • Methodologies used are under review[209]	• Recognised as weak in UNIFEM meta-evaluation; noted that measuring impact would require participatory and innovative approaches and more resources • Outcome mapping[210] • Impact evaluation strategy being implemented in a global programme[211]

Several UN entities advocate impact measurement in their policies and various guidelines: for example, according to a UNDP handbook, impact evaluations are useful when the 'project or programme is functioning long enough to have visible effects' and 'has a scale that justifies a more thorough evaluation'.[212] For UN WOMEN, demonstrating evidence of impact is considered to be the primary goal of M&E activities at the thematic level, especially regarding progress towards the goals of the Fund for Gender Equality.[213] According to UNHCR's evaluation policy, the 'primary concern of all evaluations is the impact of UNHCR's work on the rights and welfare of beneficiaries, even when the evaluation relates to entities and functions that do not have a direct impact on people of concern to the organization'.[214]

Approaches to measuring impact vary significantly across the entities examined. Methods that support attribution have at times been promoted but are rarely used by the UN in practice. UNICEF's equity-focused evaluation resource centre presents a list of seven basic impact evaluation designs that include the use of randomized control trials or quasi-experimental designs, the latter using comparison groups that may or may not be statistically matched.[215] Where experimental designs are not

feasible, non-experimental designs are also advocated. [216] In practice however, the majority of UNICEF evaluations examined do not use rigorous scientific-experimental approaches.[217]

A common theme among UN actors is the fact that practical evaluations at the impact level are less abundant than other types of evaluations even while the need to measure impact is often recognised in policy guidance: for example, although little public information is available regarding DPKO approaches to impact assessment, a recent review of DPKO arrangements for M&E of SSR suggests that mechanisms for measuring impact are generally lacking.[218] This is considered to be due to a focus on measuring internal performance instead of external impact on the delivery of security and access to justice.[219] In addition to lacking the capacities or mandates for measuring impact, the actual challenges of completing evaluations have also been raised. For instance, a UNIFEM meta-evaluation noted that evaluations often maintain 'impact was not possible to assess due to insufficient passage of time, lack of baselines and/or problems inferring causality'.[220]

Other evaluations refer to impact, but the methodological basis for doing so is questionable: for example, the term 'impact' might be used, but not in accordance with the OECD DAC definition. It is therefore noted that 'some terms of reference and reports use the term "impact" loosely or incorrectly'.[221] Similarly, a meta-evaluation conducted in 2005 of over 60 UNFPA evaluations noted that they tended to focus on the output level and therefore did not cover the criteria of impact and sustainability well.[222] This is echoed in a 2008 UNODC report that noted that there has been 'too much emphasis on process rather than impact'.[223] A global UNIFEM meta-evaluation also notes that neglecting impact is a key weakness in UNIFEM's approach to evaluations. It notes that the evaluation of impact should be more consistently applied (alongside the other OECD DAC criteria) when conditions are favourable, and that the assessment of impact would 'require the use of participatory and innovative evaluation approaches – which in turn would typically require additional resources for evaluations'.[224] The only DPKO impact assessment found publicly available was commissioned by its gender unit to examine the impact of the implementation of UN Security Council Resolution 1325 in peacekeeping. It used a participatory approach based on the perceptions of contribution

identified by government and civil society respondents, as well as comparison with pre-mission conditions.[225]

In the humanitarian world there has been increasing interest in the idea of adopting a joint impact assessment approach for interventions. To this end an Inter-Agency Consultation Group was created to define such an approach.[226] This was based on consultations with a wide group of actors to identify what the best approaches for measuring impact would be. The consultations were based on interviews with affected communities in 'Sudan, Bangladesh and Haiti, and local government and local NGOs in the same countries; with national government and international humanitarian actors in Haiti and Bangladesh; and with 67 international humanitarian actors, donors and evaluators in New York, Rome, Geneva, London and Washington'.[227] While this group was specifically looking at humanitarian interventions, it is of relevance given the inclusive consultation process it used as a basis for its development of a joint approach to evaluation.

A key element that came out of the process was the lack of support for quasi-experimental design because of both ethical and practical challenges.[228] While it was noted that quasi-experimental design is the preferable approach when seeking to determine attribution, it was recognised that 'the trade-offs between appropriateness and rigour need to be taken into consideration. Quasi-experimental design, usually involving large-scale surveys where enumerators have limited interaction with affected people but rather fill out pre-set questionnaires, is not conducive to participation.'[229] In terms of the approaches that were identified as most useful for measuring impact in the field, these were mainly participatory. Goal-free evaluation was also raised as being an important approach due to its participatory nature and ability to capture unintended and indirect results.[230]

Summary

While the information from this section is not specifically linked to rule of law and security institutions issues (due to the challenge of finding impact assessments focusing on these issues), there are several insights that can be drawn that are nonetheless relevant to this sector.

First, it is evident that all the actors examined are struggling with the same need to measure impact. For the bilateral actors this is sometimes a

result of budgetary and oversight requirements set out by their respective governments. For UN entities it is often related to the increased recognition of the importance of measuring impact from a policy perspective as well as pressure from member states to justify the use of resources. [231] For multilateral actors, the steps towards measuring impact can be inscribed within the efforts of the international networks they belong to that promote impact evaluation. While in most cases attribution is promoted due to the perception that it is a more rigorous approach, in practice demonstrating contribution is often seen as more feasible. However, even examples of impact measurement through contribution are limited.[232] In fact, it is often noted that more focus is needed on innovative approaches for the measurement of impact through contribution. For example, a UNIFEM meta-evaluation has promoted the use of outcome mapping and MSC approaches as options to consider and in addition to contribution analysis, multilateral and bilateral actors are also increasingly using these two techniques. Participatory approaches in particular have gained increasing recognition as a way to enhance impact measurement. [233] However together with utilisation-focussed approaches participatory approaches are also most commonly singled out in UN and non-UN actors' policy documents and handbooks as needing more attention.

Second, this brief review highlights that in cases where experimental or quasi-experimental scientific evaluation methods were employed to demonstrate attribution, these were very rarely undertaken in the area of rule of law and security institution activities: these approaches were mostly used in sectors such as rural development, education and health. This is reflected in a USAID meta-evaluation noting that other approaches were often used in the area of peace and security. The explanation put forward was that 'education, health, and economic sectors rely to a greater extent on quantitative data than do evaluations in other sectors, and are therefore more inclined to statistical/quasi-experimental designs'.[234] A similar finding is evident from the overview of the approach promoted by the OECD. In terms of evaluations in the area of development, it advocates the use of scientific-experimental impact evaluation. In the area of peacebuilding, however, the guidance recognises the utility of mixed approaches depending on the context at hand.[235]

Finally, in the few documented cases of scientific-experimental approaches being used in the area of peacekeeping, notable challenges

were identified. For example, one study on the impact of UNOCI on the population pointed to the challenge of there being 'no catch-all way to separate the effects of an intervention, on the one hand, from effects of the factors of an area that prompted the intervention there in the first place'.[236] As the assignment of peacekeeping presences to particular areas is based on necessity, it is difficult to construct a counterfactual comparison.[237] Despite these constraints, the authors of these impact evaluations nonetheless noted the value of such scientific-experimental approaches in measuring the impact of peacekeeping missions. However even though these evaluations considered an area closely related to support for rule of law and security institutions, they still did not focus specifically on such support and as a result are of limited relevance in understanding the challenges of measuring this kind of impact specifically. For example, the impact evaluation on UNOCI examined how the presence of the peacekeeping mission has given citizens a greater sense of security rather than how the mission's support to institution building has affected the well-being of the population while an evaluation report on the impact of UNMIL examined whether 'proximity to deployments is associated with more or less crime'.[238] Hence, while the evaluations were able to point to whether the presence of a peacekeeping mission had a direct impact on their intended beneficiaries, they did not examine the value of specific activities. Moreover, in line with the limitations posed by such methods, the evaluations were restricted in their ability to substantially explain why specific impacts were achieved or not.

This section has shown that although most international actors recognize the importance of measuring impact, there is no agreement on the most effective approach to doing so. While scientific-experimental approaches are often promoted in policy guidance as being more 'rigorous', in practice, other approaches to measuring impact have gained recognition. Moreover, these other approaches (e.g. theory-based, participatory, utilisation-focused) are increasingly considered to provide significant advantages when measuring impact in complex environments. This is because they may be more apt at answering the questions international actors are seeking to understand and because they are based on more accessible skill-sets and more modest resources.

The next section will take this analysis further by highlighting how impact can be measured in practice in peacebuilding contexts. In particular,

it will examine questions such as the level at which impact should be measured, the human and financial resource constraints that affect how impact can be measured, the use of indicators in measuring impact, and engagement among international and national actors when measuring impact.

MEASURING IMPACT: KEY ISSUES FOR PEACEBUILDING SUPPORT

There is no single best practice for measuring the impact of peacebuilding interventions on rule of law and security institutions in host countries. Rather it is important that actors and institutions adapt methodologies based on their own requirements and specificities. In doing so there is a need to understand the lessons and good practices that have already been developed by the international community. Moreover, the particularities of peacebuilding contexts that can influence the how impact can be measured need to be taken into account. This section takes the analysis set out in the previous sections further reflecting on the key issues that international actors often grapple with when considering impact measurement. The discussion looks at several key questions from a peacebuilding perspective in order provide the building blocks of an individualised approach to measuring impact.

The key questions that focus the discussion are:

1.   What to measure and when?
2.   How to measure impact in practice?
3.   What role is there for indicators in measuring impact?
4.   Which actors should be engaged, how and why?

What to measure and when?

There is often a lack of clarity within the literature on what should be measured and when. This relates to basic theoretical disagreements on the level of the results chain at which impact is located. Similarly, the fact that many international actors need to measure support efforts on a yearly basis has led to an assumption that impact should also be measured on this basis. This sub-section attempts to clarify some of these concerns.

What to measure?

There is often confusion among international actors between measuring outcomes and impact.[239] Impact should be measured in accordance with the definition outlined in the OECD DAC guidance. This means focussing on impact at the level of goals while outcome is located at the lower level of the objectives of an intervention. Impact is widely considered more difficult to measure than outcomes because it is harder to show the connection between the intervention and the effect. That is why in practice a majority of evaluations are focused at lower levels of the results chain. If the aim is to measure impact, then outcomes will nonetheless be measured, but efforts have to be made to go beyond measuring outcomes in order to attempt to demonstrate how an outcome has led to an impact.

A certain realism is required as to what it is possible to measure and how this measurement squares with the potential impact a project can achieve: that is to say, it would be unfair to expect one small project in the area of rule of law and security institutions to have an impact at the level of enhanced security and justice for all. This also relates to the question of which level an impact is being measured at, whether the institutional level, the beneficiary level, or some other level? For the development community, impact is primarily focused at the beneficiary level.[240] Tracing impact at this level is still likely to involve measuring impact at the institutional level, but would have the added value of looking beyond that. This approach is reflected in the UNDP's evaluation handbook, which notes that an impact evaluation 'includes the full range of impacts at all levels of the results chain, including ripple effects on families, households and communities; on institutional, technical or social systems; and on the environment'.[241] Therefore, rather than being limited to evaluating direct

effects on intended beneficiaries, it would look at effects on a second layer of beneficiaries. A similar understanding is reflected in the area of peacebuilding: the OECD DAC proposes a definition of impact for peacebuilding contexts which reads as 'the results or effects of any conflict prevention and peacebuilding intervention that lie beyond its immediate programme activities or sphere and constitute broader changes related to the conflict'.[242] This definition would meet needs in the area of rule of law and security institutions, which entails that impact should be measured in terms of behaviour and attitudinal change as well as institutional change.[243]

Another related issue is that of who should define the impact to be measured? That is to say, there are arguments that national actors should define the impact they are striving to achieve while in practice it is often international actors or donors that set the tone of the evaluation. The question is relevant because the impact that an international actor sets out to achieve may not correspond to stakeholders' expectations.[244] Identifying criteria for success in collaboration with national stakeholders would have the advantage of enhancing the relevance and context-specificity of programmes. This could be supported through participatory or action-oriented approaches determining what is to be measured on a case-by-case basis, that is, essentially collectively defining the criteria for success. This requires, however, flexibility in being able to accept potential deviance from the originally intended goals.

When should impact be measured?

There is a tendency for impact to be perceived as so long-term that there is a need to wait years to measure impact. However, if one takes the definition of impact to include both longer-term and shorter-term impact, then evaluations at the impact level can also take place relatively soon after and even within the lifespan of an intervention. In practice, many international actors must report on their progress towards achieving expected accomplishments on a yearly basis. However this should not be confused with the need to measure impact. Impact assessments do not need to take place on a yearly basis, as they are complex undertakings that may require some time to have passed in order to show visible effects. Moreover, even within the same intervention, there may be certain components that lead to long-term impact while others lead to short-term

impact. These issues raise the question of when and with what frequency international actors should be attempting to measure impact.

Options for when to measure impact include starting points both before and after the end of an intervention. Measuring impact ex-post would entail the evaluation taking place at the end of the intervention or even several years later. This is less useful if the purpose of measuring impact is to learn and especially to allow adjustments in programming during the intervention.

An alternative option is therefore to measure impact during the intervention. This can include several scenarios. The first would be to measure impact at the terminal phase (e.g. in the end phase of an intervention but before completion). This would have the advantage of giving more time for impacts to occur while still allowing for minimal adjustments. Furthermore, the lessons identified can be fed into a final evaluation report that will improve future policy and planning. The second option would be to measure impact within a broader evaluation on a yearly basis, but only including impact as one component to be measured alongside other OECD DAC criteria. However, based on the overview of experiences of international actors, this risks watering down efforts to specifically focus on impact as in these kinds of approaches specific methodologies for measuring impact are rarely used. In fact, impact is often addressed as an 'add-on' in these evaluations. A third option would be to measure impact on a yearly basis, but restricting the assessment to smaller, more limited elements, which cumulatively support a final more general impact assessment: for example, in the case of contribution analysis this could involve selecting a few strands of theory of change to test per year. Finally, a last option would be to conduct impact assessments on an ad hoc basis, for instance every three to five years depending on the needs of the intervention. This latter approach would provide the most flexibility for international actors, while at the same time ensuring that impact is being measured when there is really a need. This may best satisfy cost-efficiency concerns however it also means giving up the potential advantages of strategically planning for evaluation as an integral part of an intervention and could lead to poorly planned and hastily executed assessments.

Ideally, there should be clear criteria in place on when impact should be measured. Guidance on impact evaluation suggests assessments at the impact level should be undertaken when: the intervention has taken place

for enough time to show visible effects; the scale of the intervention in terms of numbers and cost is sufficient to justify a detailed evaluation; and/or the evaluation can contribute to 'new knowledge' on what works and what does not work.[245]

How to measure impact in practice?

While international actors have used a variety of approaches and methodologies for measuring impact, there is no general agreement on how best to broach this question in practice. Common questions include: whether to choose an approach that enables attribution or contribution? Which approach or methodology to select? How to factor in the time, human and financial resource considerations? This sub-section looks at these questions in detail in order to highlight the advantages and disadvantages of the different approaches from a peacebuilding perspective.

Attribution or contribution?

When it comes to measuring impact, there is much debate between the value of measuring attribution or evaluating contribution. There is no right or wrong answer and essentially the answer to this question should depend on a range of factors, including the purpose of the evaluation, cost-effectiveness, and peacebuilding concerns, each of which is discussed below.

- *Purpose of the evaluation:* If the evaluation is being conducted solely for accountability reasons then 'proving' impact on the basis of attribution may be more important than seeking to understand the contribution of an intervention. If the purpose is to learn from the intervention in order to 'improve' support efforts then approaches that enable contribution may be more adequate as they can capture the bigger picture by identifying the factors that affected the success of the intervention.[246]

- *Evaluation questions:* If the evaluation questions centre on quantifying impact (e.g. How much impact has been achieved? What difference has this intervention made?), then the scientific-

experimental approach that shows attribution may be useful. However, if the purpose is to answer evaluation questions aimed at learning (e.g. What went right/wrong? Why did this invention work or not work?), then methodologies supporting contribution would be more suitable (except in the case of theory-based impact evaluation which can answer both sets of questions).

- *Balancing cost-effectiveness:* Cost-effectiveness needs to be considered based on the purpose of the impact assessment. Attribution approaches are generally more expensive than measuring contribution and are likely to require greater justification for their undertaking.

- *Peacebuilding constraints:* In most post-conflict contexts there are multiple actors working in the same sector making attribution harder to demonstrate. There are also ethical and logistical questions surrounding the selection of target groups to receive benefits.

- *Considerations of national ownership:* Seeking to prove attribution can undermine national ownership in that there is an attempt to claim success through international interventions, thus minimising the effects of national engagement towards national goals.

Table 7 recapitulates some of the advantages and disadvantages of attribution versus contribution as discussed in previous sections:

Table 7:   Advantages and Disadvantages of Attribution versus Contribution

Attribution		Contribution	
Advantages	**Disadvantages**	**Advantages**	**Disadvantages**
• Develops quantitative proof of an intervention's impact	• May undermine national ownership • Less suited for considering the role of other international actors • Ethical questions • Less suited for complex, multi-site, integrated programmes • Costly • Requires highly specialised skill-set	• Demonstrates plausible contribution to an impact • Supports learning by understanding what worked/did not work and why • Generates stronger understanding of various roles of international and national stakeholders • Less costly	• Cannot quantify impact

Which methodological approaches should be adopted?

The choice of methodological approach is dictated by the decision to establish attribution or to demonstrate contribution (see discussion above). If attribution is to be established, the scientific-experimental approach must be adopted, and can be combined to a certain extent with other approaches such as theory-based evaluation. If contribution is to be demonstrated, options include combining elements of participatory, theory-based or results-based approaches among others.

Combining participatory and theory-based approaches may be the most appropriate approach for adapting to the complex environments typical of peacebuilding. Several of the methodologies identified in section 2 combine participatory and theory-based approaches (e.g. outcome mapping, contribution analysis). In terms of the scientific-experimental approach, there is an emerging consensus among the evaluation community that experimental and quasi-experimental methods 'are useful only for discrete and relatively simple interventions geared to precise and

measurable objectives and characterized by well-defined "treatments" that remain constant throughout program implementation'.[247] A further option is to use these methods to answer carefully limited and well defined questions within a broader qualitative evaluation.

When thinking about approaches to use it is helpful to consider those currently being promoted by other international actors. Compatibility of approaches is useful in pursuing joint assessments. A review of international actors' policy and practice suggests that the approaches that are gaining increasing recognition are participatory and utilisation-focused evaluation approaches. This is reflected in the major consultation undertaken in the humanitarian community on how it should approach joint impact measurement. Clear preference was given to participatory approaches as well as potentially goal-free evaluation, due to the participatory nature of these methods and their ability to capture unintended and indirect results. Scientific-experimental approaches were deemed to be too logistically and ethically challenging. Utilisation-focused approaches have also drawn increased attention due to their ability to enhance the ownership of the evaluation process by the project team, and their flexibility in adapting to diverse evaluation needs.

Table 8 recapitulates some of the advantages and disadvantages addressed in previous sections:

Table 8: Advantages and Disadvantages of Different Methodological Approaches

Approach	Advantages	Disadvantages
Scientific-experimental	• Enables attribution	• Costly • Time-intensive • Requires very specialised skill-sets • Ethical challenges • Difficulties in identifying a counterfactual
Theory-based	• Supports contribution analysis • Supports identification of unintended side-effects • Supports understanding of whether implementation was poor or if the theory of change was flawed	• Medium cost • Can be complex to use in large-scale programmes with multiple theories of change
Participatory	• Supports contribution • Supports national capacity building • Supports broader ownership of evaluation and encourages follow-up action • Uncovers unintended and indirect impacts • Relatively inexpensive (depending on extent of participation)	• Subjective (although can be made more rigorous by combining different methods) • Risk of bias and conflict of interest among national stakeholders
Action-orientated	• Enables consensus on goals to be measured • Of relevance if broad goals found in project documents need to be broken down into smaller goals that can be monitored and evaluated	• Does not enable either attribution or contribution • Requires a significant change in mind-set to enable a deviation from original goals • Less useful as a data collection method, more as a tool for collectively identifying criteria of success
Goal-free	• Uncovers unintended and indirect impacts • Minimises bias	• Quite costly • Time-intensive • Results may not be concrete enough to be used for learning and accountability purposes

Results-based	• Easy to use • Part of most international approaches to evaluation	• Does not support either attribution or contribution • Medium cost (depending on number of indicators defined) • Can be difficult to adapt to changing circumstances (static) • Difficulty in establishing adequate indicators for measuring impact • Difficulty in establishing verifiable data sources • Risk of neglecting the 'how' and 'why' by only focusing on results
Utilisation-based	• Offers the most leeway for compatibility with international actors as the methodologies selected depend on the purpose of the evaluation	• Depends on the methodologies selected.

Which methodologies/methods could be considered?

There is no common agreement among international actors on the best approach to measuring impact. This is because there is no single right way to measure impact. In fact, it may be necessary to combine several methodologies and their associated techniques in order to overcome weaknesses and build on the strengths of individual methodologies. Once there is an agreement on the broad approach, specific methodologies can be considered.

Issues to consider when selecting a methodology include the challenges of conducting evaluations in peacebuilding environments. Significant material has been written on these challenges, so this paper does not dwell on them,[248] but rather sets out how some of the challenges may affect the selection of methodology.

- *Difficulty in conducting data collection.* Common challenges include security conditions hampering the ability to travel to certain sites in order to undertake data collection, baseline information not being available and data being hard to find. This speaks against methods relying on baselines (e.g. impact evaluation using experimental

design) and speaks for techniques that do not necessarily require a baseline (e.g. MSC technique or non experimental designs).

- *Importance of unintended and negative effects.* In peacebuilding settings it is of central importance to identify unintended and negative effects due to the nonlinear conditions and constantly evolving political and security situations, all of which can impact how interventions are implemented in practice. Close attention must be paid to how the intervention affects those it is intended to support, as well as other vulnerable communities. Methodologies that address this need and are therefore better suited to peacebuilding contexts include contribution analysis, MSC and outcome mapping.

- *Financial resource challenges.* For actors operating in peacebuilding environments, the issue of cost is often of great importance. This relates to, on the one hand, the evaluation approach selected (i.e. if it relies on extensive and time-consuming data collection) and, on the other hand, the skill-sets needed to collect the data (i.e. the number of evaluators and their cost). As noted elsewhere, impact evaluation methodologies based on scientific-experimental approaches are often the most expensive and alternative methodologies are often cheaper to use.

- *Human resource challenges.* Some of the methodologies discussed take up a lot of staff time, such as those techniques that rely mainly on participatory methods or that require constant monitoring as for example with outcome mapping if applied internally (although this approach can also be adapted for use as an external assessment tool). The feasibility of using certain methodologies where there may or may not be resistance from staff in adapting to new approaches also needs to be considered. For instance, contribution analysis is fairly easy to embrace because it is based on the logical framework used in results-based approaches, and simply requires some additional information be collected during the planning phases and tested during the evaluation.

- *Level of complexity.* In complex environments such as post-conflict peacebuilding contexts, there are likely to be different theory of change strands with various interrelationships that make up a

general theory of change for a sector in the country in question. Contribution analysis is particularly well suited for such complex interventions. Individual strands can be tested at different points in time to enable an iterative contribution analysis.[249] For instance, feedback could be provided on a yearly basis on some of the individual strands comprising the global theory of change.[250] This would enable rapid adjustments where needed on the individual strands while feeding into a comprehensive impact assessment that brings all the strands together. It would also allow for differential timing of individual evaluations in the recognition that some elements of the theory of change need more time to produce effects than others.

There are also characteristics of rule of law and security institutions that make some methodologies more suitable than others.

- *Political nature of rule of law and security.* SSR programmes often have vague, hidden, ambiguous and highly ambitious goals and objectives, which make performance measurement or effective results-based evaluation very challenging, especially at the impact level.[251] There are various reasons for this among which is the highly political nature of SSR and the fact that politically sensitive objectives might not be openly stated.[252] In order to conduct an evaluation of programmes with vague and hidden goals, goal-free evaluations could prove useful, since they do not look at the programme logic and its stated objectives, but at the actual outcomes and impacts.

- *The need to measure long-term impact such as behavioural change.* It is often difficult to measure tangible outcomes and a lot of support is geared towards achieving long-term change: for example, much international support is about facilitation, coordination and capacity building. This is less the case for mine action and DDR, but is particularly true for SSR and rule of law support. Outcome mapping may be well suited to measuring this kind of change since it defines impact as changes in the behaviour of key partners and stakeholders.

- *Multiple linkages between interventions within the security and rule of law sector.* There may be multiple causal chains that need to be

examined, often beyond the intervention itself: For example, activities in the area of DDR may affect the way that support to SSR is provided and its results. Contribution analysis may be the most adequate approach for dealing with this type of challenge, as it enables the examination of multiple theories of change with different contribution strands that can be tested at different moments of the lifespan of the intervention.

- *Strengthening national ownership and capacity building*. These are core principles of most rule of law and security institutions components: in the area of SSR, for instance, it has been recognised that there is 'an urgent need to broaden the range of voices heard in most evaluations so that they more accurately represent the views of the populations that are meant to be the ultimate beneficiaries of SSR programmes.'[253] Participatory methodologies such as contribution analysis, MSC and outcome mapping would therefore be useful in these contexts.

This discussion has highlighted that each methodology has its strengths and weaknesses for measuring certain elements of support to rule of law and security institutions, and they can be combined to create a more complete picture of impact (see table 9 below on the advantages and disadvantages of the various methodologies). For example, ROA is suited to looking at projects that involve support to policy changes. Outcome mapping is better suited for those projects that focus on enabling behavioural change and where outcomes may be unpredictable. When measuring the impact on a sector such as rule of law and security institutions not every project will be able to be covered. Rather, certain projects/programmes are likely to be selected for assessment to provide a snapshot of the bigger picture. The selection of interventions to be assessed is likely to depend on whether the cost of the project merits special examination, whether the project is likely to generate important lessons learned, or whether the contribution to impact is questionable.[254] On this basis different methodologies may be best thought of as a set of tools to be drawn upon for measuring impact according to their appropriateness for the interventions selected for assessment.

Table 9: Advantages and Disadvantages of Impact Assessment Methodologies

Methodology	Advantages	Disadvantages
Impact evaluation	• Enables attribution • Enables quantification of impact and provides statements for donors	• Establishing control groups may be challenging • Sufficient baseline data may be missing • Cost can be 'significant' • Requires sophisticated skill-sets (limits availability of consultants)
Theory-based impact evaluation	• Combines quantitative methods with qualitative methods that enable an understanding of why impact has or has not occurred in addition to supporting attribution	• The same challenges as faced by impact evaluation
Contribution analysis	• Enables contribution • Enables identification of unintended and negative impacts • Provides opportunity to learn	• Does not support attribution
Outcome mapping (OM)	• Supports contribution • Useful in situations where outcomes are unpredictable • Flexible and adaptable	• If used as originally intended, can be time-consuming and relies on an open organisational culture; but can be adapted for use in a less 'invasive' manner as a tool for external consultants • Can be used to evaluate projects or programmes, but these must be specific enough to enable the identification of groups that will change[255]
RAPID outcome assessment (ROA)	• Useful in assessing contribution to policy changes (through use of episode studies) • Flexible and adaptable	• Can be time-consuming and relies on an open organisational culture.

Most significant change (MSC)	• Useful in situations where outcomes are unpredictable • Identifies how change has come about and what change has been the most important • Supports participatory identification of impact and provides capacity building • Can be used to supplement other techniques in order to highlight change	• Does not measure attribution but can support contribution • Requires an open organisational culture • Less useful when there are well-defined and measurable outcomes • Time-consuming when used as originally intended • Requires combination with other methodologies/approaches

The selection of methodology will depend on the priorities of the international actor (focus on accountability versus learning, cost versus benefits, etc.). Moreover, the ideal approach would be to triangulate results by using several methods and indeed there are commonalities among several of the methodologies, which facilitate triangulation. Some commonalities include: the use of baselines, theory of change, indicators or progress markers, change stories, workshops, and qualitative data collection methods (e.g. focus groups, interviews, observation). There are also common principles, which should be used to guide impact assessments: for example, inclusiveness, testing of the underlying theory of change, mix-methods approaches using both qualitative and quantitative methods, and attention to unexpected impacts.[256]

How to factor in time, human and financial resource considerations?

Several evaluations point to the prohibitive costs of measuring impact – primarily due to the challenges of data collection for impact as opposed to performance.[257] This section provides an overview of some of the human and financial resource considerations (table 10) and then highlights how methodologies can be made less expensive and time-consuming.

Table 10: Time/Cost/Skills[258]

Methodology	Time	Cost	Skills
Impact evaluation (IE)	• Time-consuming to conduct[259]	• Very expensive[260]	• Requires an evaluation expert with quantitative and qualitative evaluation skills
Theory-based impact evaluation	• Similar to impact evaluation	• Similar to impact evaluation	• Requires an evaluation expert with quantitative and qualitative evaluation skills
Contribution analysis	• N/A	• N/A	• Requires an evaluation expert with qualitative skills, familiar with the methodology
Outcome mapping (OM)	• Time-consuming: the whole iterative process requires a 2–3 day workshop plus several sessions at the different stages[261] • Can be conducted internally or by an external expert	• Low expense if OM process conducted internally [262] • If evaluation is conducted by external evaluators, an example of cost of evaluation is US$46,000[263]	• Requires an evaluation expert or external facilitator familiar with the methodology • Qualitative research methods
RAPID outcome assessment (ROA)	• N/A	• N/A	• Requires an evaluation expert or external facilitator familiar with the methodology

Most significant change (MSC)	• Time-consuming: requires workshop or internal training on use of technique, plus running several cycles of selection of significant changes and feedback in order to reveal its full potential[264] • Can be modified for integration into an external evaluation as part of a questionnaire	• N/A	• Requires external facilitator familiar with methodology but is relatively easy to implement

In practice, several of the methodologies can be adapted so that they are less time-intensive or cost-intensive: for example, while impact evaluation based on genuine experimental design is promoted as the most rigorous measurement of impact, the situations where it can feasibly be used are limited. In fact, a 2006 study estimates that randomised control trials have only been used in 1-2 per cent of impact evaluations, approximately less than 25 per cent use baseline surveys, and robust quasi-experimental designs are adopted in less than 10 per cent.[265] There have therefore been broad attempts to make impact evaluation more adaptable to "real world" conditions, with options of techniques identified that can reduce time and costs.[266] One technique, which could be relevant for international actors is *post-intervention project and comparison groups with no baseline data*. This design uses the post-intervention group as the counterfactual based on the assumption that any differences between the pre-intervention group and post-intervention group are due to the effects of the intervention.[267] By eliminating baseline data collection, such evaluations cost significantly less than comparable studies using experimental or quasi-experimental designs.[268] However, the disadvantage with such simplified evaluation designs is that the results will be less reliable and robust from a methodological and statistical perspective.[269]

The participatory methodologies that enable capacity building of national stakeholders (e.g. outcome mapping and MSC) can also be adapted to be less time consuming by relying on external evaluation approaches as opposed to highly participatory ones. For example, the approach taken by MSC of asking beneficiaries several questions regarding the MSC can be adapted for use by external evaluators in their interviews or focus group meetings, without requiring training to be provided to beneficiaries to collect data in an iterative manner. The same goes for outcome evaluation, which can be adapted to be used at the end of an intervention by translating the intervention documents into behavioural terms to support the identification of boundary partners to be interviewed and outcome challenges to be tested.[270] In the case of a Sida evaluation, for example, external evaluators adapted the methodology to use progress markers in their surveys/questionnaires rather than relying on the generation of data by stakeholders.[271]

While this may not be the way the methodologies were intended to be used, in cases where there is a need to be pragmatic about what is feasible and adapting methodologies can still strengthen traditional evaluation approaches to focus more on impact.

What role for indicators in measuring impact?

Indicators are at the heart of the monitoring and evaluation approaches commonly used by international actors, in particular as part of the results-based management framework. Against this background, some examination is required of how indicators can be used at the impact level. While indicators should be perceived as part of a methodology for measuring impact, this question is addressed separately due to the importance often granted to indicators in M&E systems. Nevertheless it must be emphasised that indicators are not adequate tools for measuring impact on their own. They can support efforts to understand impact, but they cannot replace a robust methodology for measuring impact.

How can indicators measure impact?

Impact indicators are part of an organisation's performance measurement system.[272] They are located at the final goal level, and aim to measure

highest-level changes in the results chain. Indicators at the impact level have been used by several organisations. The usage of the term 'impact indicators' differs among them: for example, the UNDP lists three levels of indicators, impact, outcome and output.[273] In contrast, EuropeAid lists three different levels of impact indicators (in addition to regular output and outcome indicators): specific impact, intermediate impact and global impact.[274] Specific and intermediate impacts both cover the OECD standard definition of positive and negative, primary and secondary, direct or indirect, intended or unintended long-term effects produced by a development intervention, with the slight difference that intermediate impacts are longer term in nature and the last cause-and-effect chain level that can be monitored effectively and at the same time demonstrate sufficient attribution.[275] For the health sector, for example, impact indicators would include data on child mortality, life expectancy, etc. However, EuropeAid does not develop indicators for the global impact level, which corresponds to the highest-order goals, such as poverty reduction, as it considers that attribution and reasonable cause-effect linkages are impossible to establish at this level.[276]

Finally a third example can be drawn from the field of DDR**.** The IDDRS Module 3.50 on Monitoring and Evaluating DDR Programmes classifies indicators another way.[277] The two broad categories of indicators are performance and impact. A third category is proxy indicators, which can also be related to impact. Accordingly, *performance indicators* seek to measure outputs and outcomes.[278] *Impact indicators* measure the 'overall changes in the environment that DDR aims to influence'. It is noted that 'impact indicators often use a composite set (or group) of indicators, each of which provides information on the size, sustainability and consequences of a change brought about by a DDR intervention. Such indicators can include both quantitative variables (e.g., change in homicide levels or incidence of violence) or qualitative variables (e.g., behavioural change among reintegrated ex-combatants, social cohesion, etc.).'[279] *Proxy indicators* can be used in situations where data for impact measurement cannot be reliably collected, or are completely missing. They 'are variables that substitute for others that are difficult to measure directly [and] may reveal performance trends and make managers aware of potential problems or areas of success'.[280]

UNDP/BCPR also discusses proxy indicators in its 'How To Guide' for DDR, and notes that indirect or proxy indicators are typically used in cases when 'a result is an abstract concept. For example, a DDR programme measures impact ("security situation improved") by using four proxy indicators (violence, confiscated ammunition, confiscated weapons, suspects detained). Indirect indicators are also used if data for direct indicators is not collected frequently enough (e.g. household surveys are often conducted only every five years), or if data for direct indicators is too difficult, dangerous or expensive to collect.'[281]

In sum, while there are challenges to the development of impact indicators, experience shows that these challenges can be overcome in a number of satisfactory ways. A solution adopted by several actors is to divide impact indicators into those that can be realistically measured and those that would be more challenging. For example, one option would be to develop a list of intermediate indicators, which can be feasibly measured, and a list of longer-term indicators, which might not be measurable in the immediate term but can serve to track progress over a longer time span. Another interesting approach is that of developing proxy indicators as an additional category to support the capture of some of the more intangible elements of impact at the highest level. Progress on the intermediate and proxy indicators would therefore suggest progress in reaching the higher level of impact.

What are the considerations when developing impact indicators in the area of peacebuilding support to rule of law and security institutions?

In peacebuilding contexts, many practitioners feel that it is extremely difficult – if not impossible – to define indicators at the impact level.[282] Challenges in this task can be linked to two main themes: the difficulty of measuring results at such a high level, and the fact that the indicators are weakly designed to measure impact from the planning stage.

First, the issue of measuring the intangible is amplified in the area of rule of law and security institutions because issues such as 'peace, communal trust, and good governance are intangibles.'[283] It is therefore considered that peacebuilding interventions cannot be compared to development interventions in the field of health or education. This is particularly the case in the area of rule of law and security institutions

where broad goals are often affected by a plethora of factors, for example at the political and social levels. Furthermore, indicators at the impact level are often of only limited usefulness because an intervention's contribution to impact is likely to be very small and abstract: 'The project might, for example, contribute to an increase in a democracy indicator by three points, but what does this really tell the project leadership?'[284]

Second, there is a sense among practitioners that even in circumstances where indicators at the impact level are identified, they are not really adequate representations of the impact level. For example, 'number of soldiers demobilised' has been used as an impact indicator but it is not able to demonstrate improvement to beneficiaries' lives. It could more adequately be considered an output indicator.[285] This is linked to a broader challenge with the use of indicators in general in so far as there are often 'design flaws'.[286] Such flaws have been reflected in many reports of evaluations: for example, UNIFEM's 2009 meta-evaluation notes that 'the quality of the indicators in the logical frameworks varies considerably' and in some cases 'the indicators are too specific so that regular reports contain frequent repetitions of activities having been achieved. In others the indicators contain broad, unqualified statements.'[287] This problem is also mentioned in individual evaluations, which have recognised that indicators are often not tracked in practice and the right indicators are not always selected.[288] Challenges in developing appropriate indicators and using them in a regular and effective way are linked to broader difficulties in using the results-based management framework in peacebuilding contexts.

In the area of measuring performance of law and justice institutions specifically, the World Bank has advocated for the use of a "mix of indicators".[289] This appears to be even more important when attempting to measure impact, as elements of impact may be difficult to capture and therefore must rely on a mix of indicators to adequately track progress. Impact indicators need to be specifically developed at the design stage to ensure that they enable measurement at the highest level of the results chain and are not relegated to the output or outcome levels.

Finally and as discussed above, an essential consideration in the area of peacebuilding support to rule of law and security institutions is the need to be able to capture behavioural change. This is an important component of effects at the impact level and indicators are often challenged in this respect. The progress markers discussed in the context of the outcome

mapping methodology provide an interesting example as they could be used to supplement regular indicators that are not as capable of assessing behavioural change. Progress markers are not as rigorous as regular indicators in that they are not intended to follow all of the recognized criteria promoting specific, measurable, achievable, relevant and time-bound (SMART) indicators. However, they can capture changes in behaviour and relationships that are difficult to perceive with the use of more traditional SMART indicators. In particular, they enable telling the 'story of change' as opposed to providing 'one-off snapshots of change'.[290]

Who should be engaged, how and why?

While effective and sustainable peacebuilding efforts rely on international actors coming together to provide support in a coherent manner, their roles and responsibilities in this area are highly fragmented. In an ideal world, a sector-wide approach to measuring impact would be promoted to help address this challenge, as well as to measure the extent to which the support followed an integrated approach. Moreover, national actors would be engaged in order to support national ownership and capacity-building in the area of M&E. This sub-section will examine the question of who should be involved in measuring impact within the sector, what type of engagement should be promoted among international actors, and how engagement with national actors can be promoted when measuring impact.

Who should be involved in measuring impact within the sector?

Evaluating support to rule of law and security institutions can be done in a sector-wide or component-based approach. Sector-wide means that the whole sector is examined, while a component-based approach would focus on individual evaluations, for example, for mine action, rule of law, DDR etc. The rationale for sector-wide evaluations can be linked to the challenges of attributing impacts to any single project or programme, given the number of actors operating in post-conflict environments. Focusing on these higher levels therefore allows examination of the contribution of different actors to the impact observed on national institutions and beneficiaries. Moreover, sector-wide evaluation at the impact level is important in seeing the bigger picture and understanding how particular components have

complemented or undermined one another. For example, it has been noted that in the area of DDR, evaluations would benefit from being undertaken across a number of related sectors.[291] This is because DDR requires achievements to have taken place in other sectors (e.g. SSR) for it to be sustainable. Similarly, the European Commission has since 2001 centred its conflict prevention and peacebuilding strategy on an integrated approach encompassing 'different types of activity', 'different time dimensions', 'the activities of different actors' and 'different geographical dimensions'.[292] It found that while the impact of its activities had not yet fully materialised and therefore could not be assessed, the integrated nature of its approach was crucial to the positive contribution it made to mitigating the root causes of conflict in three different countries.[293]

Sector-wide approaches can sometimes be broken down into component parts. That is to say, for example, if contribution analysis was used to measure the impact at the sector level, multiple theories of change would have to be examined, each of which belongs to individual components. The final assessment would then look at how the individual strands relate to one another in order to form the 'big picture' on the sector-wide level. In the area of SSR, it has been noted that most M&E is conducted at the project or programme level – often because the project level is considered to provide a more feasible unit of analysis due to the availability of specific project documents and related logframes.[294] However, even at this component-level, there is a growing consensus 'on the need to move M&E beyond the project level to the sector and strategic level in fragile states'.[295]

The selection of approach has different implications. Sector-wide evaluations often rely on joint evaluations (see discussion below on international actors). Moreover, a sector-wide approach would imply the need for clear terms of reference and a joint understanding of criteria of success etc. This is because there needs to be a coordination mechanism at the strategic level. A number of obstacles remain to developing such an approach, including bureaucratic disincentives, different institutional cultures and timelines, and the reality that donors may prefer to account for the impact of each actor individually in order to enhance accountability over resources invested. There are nonetheless opportunities to enhance the coherence of evaluation efforts in the sector, and given the significant human and financial investments involved it is important not to reinvent

the wheel but to look for synergies and avoid duplication to the greatest extent possible. Actors may consider building on approaches that have been recognised as useful by other international actors, such as participatory and utilisation-based approaches. Simple efforts such as informing other actors of an intention to undertake a perception survey and allowing them to contribute to it or subsequently use it can reduce costs and ensure that actors are working on the basis of the same information.

What type of engagement should be envisaged among international actors?

While evaluations are often conducted by individual actors, the value of conducting joint evaluations is also gaining increasing recognition, particularly establishing impact in areas where multiple actors are engaged. When focusing on sector-wide approaches, the value of joint evaluations is further enhanced due to the challenges of differentiating between different entities' impacts in conflict prevention and peacebuilding contexts.[296] Therefore, joint evaluations have been said to 'provide the best way of assessing the cumulative and overall impacts of international programming in a single conflict context'.[297] A further advantage of joint evaluations is the ability to pool resources. For instance, to address budgetary constraints, UNICEF's meta-evaluation notes that where there is common interest, a good approach may be 'to institute a series of joint evaluations of intervention strategies in key areas with one or more partners among the international agencies'.[298]

        While there are advantages to such approaches, there are also challenges. First, if an impact assessment is conducted with the sole purpose of proving attribution to an individual actor in an effort to justify resources, there is less likely to be room for joint evaluations that look at cumulative impact. This is less of a challenge for evaluations that seek to show plausible contribution. However, in cases where the joint approach is favoured, there can still be challenges when different theories of change are used by different entities, when planning and programming timelines do not correspond, or when methodologies used are not compatible with one another (e.g. a goal-free approach would not necessarily be compatible with a theory-based approach). Finally, in the case of joint evaluations, there is also a greater need to make special efforts to strengthen local ownership given the overburdening nature of joint efforts.[299]

Compatibility of approach must be considered and embracing the OECD DAC definition of impact goes a long way towards promoting compatibility given that a majority of international actors use the OECD DAC's understanding of impact (based on intended, unintended, positive and negative consequences). Further issues to consider are other entities' need to follow certain rules and regulations. Ultimately utilisation-focused approaches for joint evaluations may leave the greatest leeway for establishing compatibility, as the approach selected depends on the specific purpose of the evaluation.

What type of engagement should international actors envisage with national actors?

International actors have increasingly recognised the need to engage with national actors when conducting evaluations. For example, a General Assembly resolution of 2004 'recognizes that national governments have primary responsibility for coordinating external assistance, including that from the United Nations system, and evaluating its impact in contributing to national priorities'.[300] This imperative must be contextualised in the growing trend towards supporting national ownership of evaluations through participatory approaches, as well as increasing joint and partner-led evaluations. For example, the OECD DAC's quality standards for development evaluation note the importance of partnership, coordination and alignment with national actors, as well as supporting capacity development.[301]

In evaluations of support to rule of law and security institutions, national actors can take on many different roles. A minimal approach would be to ensure that key stakeholders are engaged throughout the evaluation process: an example would be consulting with national authorities when developing the terms of reference for evaluation, including securing agreement on goals and theory of change. This could be extended by reaching out to those stakeholders that may be affected by the evaluation to make them aware of the evaluation and support their willingness to engage with the process. Ultimately, efforts should be made to ensure that the evaluation process recognises national and local evaluation plans and considers potential synergies.[302]

Other approaches with a higher degree of participation include involving national evaluators directly in the evaluation team itself which could mean ensuring that there are national experts on a team, or actively facilitating national actors to collect and monitor data (e.g. in the case of outcome mapping or MSC). The rationale for including national actors in such teams is simple: they have greater knowledge of local culture and politics, including stakeholders and how to engage with them, and often possess the necessary language skills. It also supports efforts to build the evaluation capacity of the country. Furthermore, it would reduce the common perception that evaluations are 'donor-centric', aiming solely to accommodate the needs of international actors. If evaluation results are not produced with support from national partners, the actual use of the evaluation findings will often be very limited.[303]

However, participatory evaluation approaches pose risks and challenges, especially in conflict and post-conflict settings. A significant challenge is the potential for biased and distorted findings when beneficiaries, partners and stakeholders have been involved in an armed conflict.[304] The culture of secrecy in rule of law and security institutions in host countries can also be a major obstacle for the collection of data and use of evaluation results.[305] The lack of sufficient evaluation expertise on the national level is another challenge often encountered[306] and one that highlights the need for increased focus on national capacity-building for evaluation. Finally, one of the most significant challenges is finding a way to accommodate the different evaluation needs and perspectives of national partners, international donors and beneficiaries.[307] Action-orientated, goal-setting approaches and utilisation-focused evaluation can prove useful in aligning all stakeholders behind common goals and objectives.

The exact degree of participation and ownership should be carefully calibrated to the specific intervention and context. The OECD Guidance on Peacebuilding Evaluations suggests some questions to consider when deciding on the extent to which local beneficiaries, stakeholders and partner governments should be included in the design and conduct of the evaluation. These include reflecting on the degree of politicisation in the country, the potential for bias in the evaluation of success based on power relations, the feasibility of supporting the collective definition of theory of change, indicators, etc., and the ability to ensure that all relevant

stakeholder views can be included in the evaluation in a meaningful manner.[308]

Summary

This section has examined some of the key concerns for international actors seeking to measure impact in peacebuilding contexts. First, it examined the level and frequency at which impact should be measured. It has highlighted that it is not necessary to measure impact on a yearly basis but should instead be measured when there is a need to learn new information on what works and what does not work or when there are doubts about the extent to which the intervention is achieving expected results. Second, this section examined the advantages and disadvantages of different evaluation approaches and methodologies from a peacebuilding perspective. In particular, the discussion highlighted that there are significant differences in terms of resources and time required to conduct an evaluation depending on the methodology selected. It was also underlined nonetheless that there are ways to adapt methodologies to suit the demands of complex contexts and limited resources thereby making them more accessible for international actors. Third, this section reflected on the role of indicators in measuring impact. In particular, it was noted that while there are numerous challenges to establishing appropriate indicators at the impact level, these challenges have been overcome in several instances by various international actors. One approach is to divide indicators at the impact level into those that can realistically be measured and those that would be more difficult to use. Finally, this section considered who should be engaged in measuring impact, how and why. It was noted that there are increasing calls to promote sector-wide approaches to measuring impact that would require joint approaches to measuring impact. There is also an increasing recognition of the need to engage national actors when measuring impact in order to support national ownership and capacity-building. In sum, this section highlights that while there are genuine challenges to measuring impact in peacebuilding contexts, it can be done. Successful impact measurement begins with an understanding of the good practices that have already been developed as well as recognition of the fact that any approach should be based on the individual needs and specificities of each actor and intervention.

CONCLUSION

Understanding impact is necessary and desirable for international actors engaged in peacebuilding interventions. Support to rule of law and security institutions is a particularly important part of this peacebuilding agenda, but although closely linked these two activities are often promoted in a disconnected manner. Measuring the impact of international support in this area can therefore help to identify synergies, build coherence, and maximise the potential positive impacts they can achieve. Focusing on peacebuilding contexts and interventions, this paper has explored a range of approaches to measuring the impact of support to rule of law and security institutions and how these approaches and methods have been applied in practice. Approaches range from those that can demonstrate attribution to those that can evaluate contribution. The specific impact assessment methodologies identified include impact evaluation, theory-based impact evaluation, contribution analysis, outcome mapping, MSC, and ROA.

While acknowledging that much remains to be learnt in this area, the study nonetheless offers some key findings. First, there is no common agreement among international actors on the best approach to measuring impact. This is because there is no 'best' approach. Each approach and methodology has its strengths and weaknesses. In fact, the best way to measure impact is to combine several approaches and methodologies in order to build on their individual strengths and mitigate their weaknesses according to the context at hand. This recognizes that the triangulation of

methods offers the most valuable approach to developing strong impact statements. Moreover, this paper argues that the selection of a combination of approaches and methodologies to be applied should be based on four criteria: firstly, the purpose of the evaluation (accountability or learning); secondly, the questions the evaluation seeks to answer (What impact? Why was the impact such?); thirdly, cost-effectiveness in view of the task at hand; and fourthly, the specific constraints of the peacebuilding context.

A second key finding relates to the scientific-experimental approach to evaluations, which has often been promoted in the development field as the only 'rigorous' approach to measuring impact. This is because it is based on counterfactual analysis – that is to say an understanding of the situation of the beneficiaries had the intervention not taken place – substantiated by control trials and subject to statistical analysis. While a counterfactual is necessary for measuring impact in the sense of attribution, this paper has also shown that impact can be measured on the basis other evaluation approaches. Theory-based approaches (i.e. testing the theory of change) and participatory approaches (looking to beneficiaries in order to hear practical examples of what has changed in their lives) can also provide the basis for establishing a suitable counterfactual. Actors seeking to measure impact can therefore adopt a host of alternative approaches that may be more amenable to the complexities of peacebuilding environments and the goals and resources of the evaluation. This entails being able to use methodologies that might have lesser costs or skill requirements than those traditionally promoted as a basis for impact measurement.

A third key finding is the understanding that there are small steps that can be taken to strengthen traditional evaluation approaches to focus more on impact. For instance, if the capacity or will to invest in fully-fledged impact assessments is lacking, it is possible nonetheless to adapt these methodologies to be less costly and less time consuming. Impact evaluation can be rendered less expensive by using alternative techniques such as post-intervention project and comparison groups with no baseline data. Participatory methodologies, which are often dismissed as being too time consuming, can also be made more user-friendly: for instance, in the case of the MSC technique, rather than relying on numerous participatory workshops and storytelling, an external evaluator can insert the 'most significant change' question into interviews or focus group meetings. This

also removes the need to provide training to beneficiaries on data collection and analysis. The same goes for outcome mapping: progress markers can be integrated into surveys rather than relying on the generation of data by stakeholders. While this is not the way these methodologies were originally developed for use, such adaptation allows international actors to focus more on impact by supplementing their traditional evaluations with new techniques.

A fourth key finding notes that measuring impact can be a significantly political undertaking. There is a risk that an evaluation may shed light on the failings of an intervention to achieve its desired impact, and in that case there needs to be clarity on whether all actors are willing to confront this reality and what they can do with this information. This requires a certain realism in terms of the expectations of both the intervention and of the evaluation. Concerning the intervention, it must be clear that a relatively small project cannot be expected to have a huge or even a significant impact. Nor is it likely that it will have an impact at the level of broader peace and security. Clarity and agreement are needed from the outset on what type of impact can be reasonably expected. Second, an evaluation cannot be both rigorous and 'quick and dirty'. That is to say that there is a risk that assessments that have not been conducted with sufficient planning, time and resources, may be placed on a pedestal and used to make important decisions simply because they were labelled impact assessments. However, a proper impact assessment requires significant resources and goes beyond a basic evaluation at the output level. Actors must understand what can realistically be expected from the range of available methodologies but also from the resources invested.

Finally, and against this background, attempting to measure impact is – or ought to be – more expensive and time-consuming that an evaluation at a lower level of the results chain. It should be clear that there is no need to measure impact on a yearly basis. In fact, it has been suggested that impact should be measured when the intervention has been in place for long enough to show observable effects, that the scale of the intervention (numbers and cost) justifies measuring impact, and/or the evaluation can contribute to new understandings on what works and what does not.[309] The same logic can be applied to what should be measured within a large intervention – not every project needs to undergo an impact assessment, freeing resources to target those areas of the theory of change which are

least clear and would benefit most from a more thorough assessment. An informed distribution of resources would help to mitigate a common critique of peacebuilding activities, which is that there is a lack of clarity as to whether an intervention is achieving its expected results due to an inconsistent theory of change or challenges in the way the intervention is being carried out.

In terms of the way ahead, there is a need to recognise that measuring impact provides an opportunity to engage with national actors. There have been increasing calls within the policy community to support national capacity-building through evaluations. Building capacity for evaluations is fundamental for the professionalism and effectiveness of security and justice institutions, and yet is an oft-neglected component of such programmes. Using elements of participatory approaches can help to build this national capacity, as well as foster an understanding among national actors of M&E as a normal and useful component of such activities. While supporting national ownership through impact assessments should be encouraged, the extent of this kind of capacity-building should be decided according to the conditions of each context.

In moving forward there is also a need to promote a 'culture of learning' as opposed to a 'culture of blame' in evaluations.[310] This is likely to resonate strongly in the case of impact assessments, which as discussed above can be highly political and politicised undertakings. International actors should take this challenge into account when developing their approach to measuring impact, and should accompany their attempts to measure impact with an effective communications strategy that highlights the importance of evaluations as a learning mechanism. Establishing a culture of learning also relates to the need to ensure that the results of an evaluation are actually used to support adjustments in policy and practice based on an enhanced understanding of what is working and what not. Procedures for ensuring this is the case should be included in any approach to measuring impact – including timelines and responsibilities. Moreover, mechanisms for sharing findings with national (and other international partners) should be developed.

This paper has highlighted that measuring impact is essential in the area of peacebuilding support to rule of law and security institutions. Moreover, it shows that there are a range of feasible approaches and methodologies to support international engagement in this area. Ultimately

the approach taken will vary from case to case. Piloting various approaches will lead to a better understanding of what works for different actors operating in quite distinct contexts. There is also a need for applied research on the practical implementation of theoretical and methodological approaches to measuring impact. This includes the compilation of detailed knowledge on evaluations that have used different techniques in the area of support for rule of law and security institutions. Only then can we grasp the quite distinct challenges encountered as well as achieve a more sophisticated understanding of the methodologies that can be best applied to learning about the impact of peacebuilding interventions on rule of law and security institutions.

NOTES

1   This paper is adapted from a mapping study on impact assessment methodologies provided for the Office of Rule of Law and Security Institutions (OROLSI) at the United Nations Department of Peacekeeping Operations (DPKO) and was first presented at a workshop convened by DPKO with the support of DCAF on 'Measuring the Impact of Peacekeeping Missions on Rule of Law and Security Institutions' on 12 March 2012. The author is grateful to Alan Bryden, Mpako Foaleng, Antoine Hanin, Heiner Hänggi, Albrecht Schnabel, Victoria Walker and an anonymous reviewer for their insightful comments on this and previous versions of the draft and to Cherry Ekins, Simon Kläntschi and Callum Watson for their valuable research and editing support. The author is grateful to have benefited from useful comments provided by experts at the DPKO workshop, and in particular to Michael Bamberger for having reviewed the study in preparation for the expert workshop. Finally, the author wishes to thank Ilene Cohn, Annika Hansen and Anna Shotton from DPKO OROLSI for their fruitful collaboration on this project. The views in this paper are those of the author alone and do not reflect the views of the institutions or experts named.

2   Decision of the Secretary-General's Policy Committee, May 2007, cited in United Nations, *UN Peacebuilding: an Orientation,* (New York: United Nations Peacebuilding Support Office, 2010), 5. Available at: http://www.un.org/en/peacebuilding/ pbso/pdf/peacebuilding_orientation.pdf

3   United Nations, *United Nations Peacekeeping Operations. Principles and Guidelines* (United Nations Department of Peacekeeping Operations / Department of Field Support, 2008), p. 26-27.

4   United Nations, *Uniting our strengths: Enhancing United Nations support for the rule of law,* Report of the Secretary-General, Document No. A/61/636–S/2006/980, 14 December 2006, para. 53.

5   United Nations, Securing peace and development: the role of the United Nations in supporting security sector reform, Report of the Secretary-General, Document No. A/62/659–S/2008/39, 23 January 2008, para. 14.

6   See footnote 1.

7   See http://www.un.org/en/peacekeeping/about/dpko/.

8   Governance and Social Development Resource Centre (GSDRC), *M&E in Fragile States*, Helpdesk Research Report, (Birmingham, UK: University of Birmingham, October 2007), p. 1.

9   See for instance recent efforts by the UN Inter-Agency Working Group on DDR to examine the linkages between DDR and SSR and DDR and Transitional Justice in the UN Integrated DDR Standards.

10 For example, DPKO OROLSI has received requests from member states to consider further ways to measure the impact of its peacekeeping components in the area of

support to rule of law and security institutions. The need to measure impact is also reflected in the policies and guidance of numerous international actors: see section 3.

[11] See discussion in: Diana Chigas and Peter Woodrow, 'Demystifying Impacts in Evaluation Practice', *New Routes* 13, no. 3 (2008): 19 – 22.

[12] Diana Chigas and Peter Woodrow, 'Demystifying Impacts in Evaluation Practice', *New Routes,* vol. 13, no. 3, autumn 2008, pp. 19 – 22.

[13] Frans Leeuw and Jos Vaessen, *Impact Evaluations and Development - NONIE Guidance on Impact Evaluation* (Washington DC: NONIE, 2009)*,* p. 47.

[14] This paper is based on a background paper prepared for DPKO-OROLSI that was presented at a workshop in New York on the 12th of March 2012.

[15] Selection criteria are described in section 3.

[16] DCAF established an internal review team that provided feedback on the study. A previous version of the study was also presented at a UN DPKO OROLSI workshop in New York where a number of experts were able to provide feedback. (see Workshop convened by DPKO in collaboration with DCAF, 'Measuring the Impact of Peacekeeping Missions on Rule of Law and Security Institutions', 12 March 2012, New York). Limitations of this study include the very short timeline available to conduct the research and draft the original study in preparation for the DPKO workshop (see footnote 1). A more systematic study covering all relevant evaluation reports could thus not be conducted within the time available. This limiting factor was mitigated by the use of meta-evaluations to provide an analysis of a broader sample of evaluation approaches, as well as examining individual reports.

[17] The Annex of the NONIE guidance note identifies a range of possible methodologies, including for example MAPP (Method for Impact Assessment of Programmes and Projects). The approaches covered in this study have been selected on the basis that they hold the most promise for measuring support to rule of law and security institutions because they consider specific constraints such as among others the complex and political nature of this kind of support and the need to measure behavioural change. See: Frans Leeuw and Jos Vaessen, *NONIE Guidance on Impact Evaluation.*

[18] OECD DAC, *Glossary of Key terms in Evaluation and Results Based Monitoring* (Paris: OECD, 2002), p. 24.

[19] OECD-DAC, *Guidance on Evaluating Conflict Prevention and Peacebuilding Activities* (Paris: OECD, 2008), p. 65.

[20] Cedric de Coning and Paul Romita, *Monitoring and Evaluation of Peace Operations*, (New York: International Peace Institute, 2009), p. 4.

[21] OECD DAC, *Glossary of Key Terms*, pp. 21-22.

[22] The International Association for Impact Assessment (IAIA) defines Impact Assessment as "the process of identifying the future consequences of a current or proposed action." See http://www.iaia.org/aboutiaia.aspx.

[23] Chris Roche, *Impact Assessment for Development Agencies: Learning to Value Change* (Oxford: Oxfam/Novib, 1999), p. 18.

24 Broad international initiatives in the past decade that highlighted the need for a global shared approach to international cooperation as well as increased aid effectiveness and better development results, include: The Millennium Development Goals, the OECD DAC Paris Declaration of Aid Effectiveness (2004) as well as the Accra High Level Forum for Aid Effectiveness (2008).

25 Kenneth Busch, *A Measure of Peace: Peace And Conflict Impact Assessment (PCIA) Of Development Projects In Conflict Zones,* Working Paper No. 1, The Peacebuilding and Reconstruction Program Initiative & The Evaluation Unit (Ottawa: International Development Research Centre, 1998). Research and guidance on PCIA were advanced in the early 2000s by the Berghof Conflict Research Center. For a collection of articles on PCIA between 2000 – 2003, see: Alex Austin, Martina Fischer and Oliver Wils (eds.), *Peace and Conflict Impact Assessment. Critical Views on Theory and Practice,* Berghof Handbook Dialogue Series 1 (Berlin: Berghof Research Center for Constructive Conflict Management, April 2003).

26 Mary B. Anderson, *Do No Harm: How Aid Can Support Peace - or War* (Boulder/London: Lynne Rienner Publishers, 1999).

27 Thania Paffenholz and Luc Reychler, *Aid for Peace – A Guide to Planning and Evaluation for Conflict Zones* (Baden-Baden: Nomos, 2007).

28 The Aid for Peace Approach is one of the first step-by-step process frameworks to guide project planning, monitoring and evaluation for peacebuilding interventions. Two other major contributions to guidance on peacebuilding evaluations include: John Paul Lederach, Reina Neufeldt and Hal Culberton, *Reflective Peacebuilding: A Planning, Monitoring and Learning Toolkit* (Notre Dame, IN/Manila, Philippines: The Joan B. Kroc Institute for International Peace Studies, University of Notre Dame/Catholic Relief Services Southeast Asia Office, 2007); Cheyanne Church and Mark M. Rogers, *Designing for Results: Integrating Monitoring And Evaluation In Conflict Transformation Programs* (Washington DC: Search for Common Ground, 2006).

29 Cheyanne Scharbatke-Church, 'Evaluating Peacebuilding: Not Yet All It Could Be', in Alex Austin, M. Fischer and H.J. Giessmann (eds.), *Advancing Conflict Transformation - The Berghof Handbook II*, (Leverkusen: Barbara Budrich Publishers, 2011), pp. 461 - 462

30 Cheyanne Church and Mark M. Rogers, *Designing for Results*, p. 103.

31 An overview of these views is described in: Cheyanne Church and Mark M. Rogers, *Designing for Results,* p. 12.

32 Christoph Spurk, 'Forget impact: concentrate on measuring outcomes: lessons from recent debates on evaluationof peacebuilding programmes', *New Routes,* vol. 13, no. 3, autumn 2008, pp. 11 – 14.

33 Isak Svensson and Erik Brattberg, 'Randomisation: A New Approach to Measure Impact of Peacebuilding Interventions', *New Routes,* vol. 13, no. 3, autumn 2008, pp. 23 – 24.

34 For example, the challenges of using impact evaluation in peacekeeping contexts were illustrated by the problems encountered in an assessment of UNOCI when the limited reintegration activities in DDRRR (early stages) hampered the ability to measure impact.

Eric Mvukiyehe and Cyrus Samii, *Laying a Foundation for Peace? A Quantitative Impact Evaluation of the United Nations Operation in Cote d'Ivoire,*, December 2008.

[35]   In the field of SSR, for example, there is increasing recognition of the fact that "security and justice programmes should be seeking to *contribute* towards positive change at the impact level, but should not expect the M&E system to *attribute* such changes to the programme". Duncan Hiscock and Simon Rynn, 'Monitoring and Evaluation', *OECD Handbook on SSR,* p. 17.

[36]   Diana Chigas and Peter Woodrow, 'Demystifying Impacts in Evaluation Practice'.

[37]   See: Mary Anderson and Lara Olson, *Confronting War: Critical Lessons for Peace Practitioners,* (Cambridge, MA: CDA, 2003).

[38]   OECD DAC, *Guidance on Evaluating Conflict Prevention and Peacebuilding Activities.*

[39]   *Ibid*. p. 42.

[40]   Cheyanne Church and Mark M. Rogers, *Designing for Results*, p. 114.

[41]   Theory of change refers to the identification of building blocks that are needed to achieve overall goals. Linkages are established between the various levels to understand what is needed to produce the next level of results. Definition adapted from sources available on the internet resource site dedicated to theory of change: http://www.theoryofchange.org/about/what-is-theory-of-change/ ;

[42]   See for instance: Frans Leeuw and Jos Vaessen, *NONIE Guidance on Impact Evaluation.*

[43]   See for example: Martin Prowse, *Aid effectiveness: the role of qualitative research in impact evaluation*, Background Note (London: Overseas Development Institute (ODI), 2007), pp. 1 – 2.

[44]   Nicola Jones, Harry Jones, Liesbet Steer and Ajoy Datta, *Improving Impact Evaluation production and use*, ODI Working Paper 300 (London: ODI, March 2009), p. 7.

[45]   Cheyanne Church and Mark M. Rogers, *Designing for Results*, p. 118.

[46]   *Ibid*. p. 119.

[47]   For general information on participatory M&E approaches see for example: Marisol Estrella (ed.), *Learning from Change: Issues and Experiences in Participatory Monitoring and Evaluation* (Ottawa: International Development Research Centre, 2000); GSDRC, *Participatory M&E and Beneficiary Feedback*, Helpdesk Research Report, (Birmingham, UK: University of Birmingham, September 2010).

[48]   Ken Menkhaus, *Impact Assessment in Post-Conflict Peacebuilding: Challenges and Future Directions,* Paper Commissioned by International Peace Alliance, Interpeace, July 2004, p. 10.

[49]   Mary Anderson quoted in: *Ibid*. p. 11.

[50]   OECD DAC, *Guidance on Evaluating Conflict Prevention and Peacebuilding Activities*, Annex 7, p. 86.

[51]   Marc Howard Ross, 'Action Evaluation in the Theory and practice of Conflict Resolution', May 2001, http://www.gmu.edu/programs/icar/pcs/Ross81PCS.htm.

[52]   Cheyanne Church and Mark M. Rogers, *Designing for Results*, p. 121.

[53]   *Ibid*. p. 116.

[54]   *Ibid.*

[55] OECD DAC, *Guidance on Evaluating Conflict Prevention and Peacebuilding Activities*, Annex 7, p. 88.

[56] Cheyanne Church and Mark M. Rogers, *Designing for Results*, p. 117; OECD DAC, *Guidance on Evaluating Conflict Prevention and Peacebuilding Activities*, Annex 7, p. 88.

[57] Judith Wilde and Suzanne Sockey, *Evaluation Handbook* (Albuquerque, NM: New Mexico Highlands University, 1995), p. 6.

[58] Michael Quinn Patton, *Utilization-focused Evaluation (U-FE) Checklist,* January 2002, p. 1. http://web.idrc.ca/uploads/user-S/10905198311Utilization_Focused_Evaluation.pdf

[59] *Ibid*.

[60] *Ibid*.

[61] *Ibid*.

[62] Danish International Development Agency (Danida), *Danida Evaluation Guidelines* (Copenhagen: Ministry of Foreign Affairs of Denmark, January 2012), p. 20.

[63] For detailed account please refer to: Frans Leeuw and Jos Vaessen, *NONIE Guidance on Impact Evaluation.*

[64] *Ibid.* p. ix.

[65] Nicola Jones et al., *Improving Impact Evaluation,* p. 4.

[66] See for example: Ester Duflo, Rachel Glennerster and Michael Kremer, *Using Randomization in Development Economics Research*: *A Toolkit*, NBER Technical Working Papers 0333 (Cambridge, MA: National Bureau of Economic Research, 2006).

[67] Eberhard Gohl, Carsta Neuenroth, Alexandra Huber and Verena Brenner, *A Guide to Assessing Our Contribution to Change* (Geneva: Acting by Churches Together for Development (ACT Development), 2009), p. 70.

[68] Frans Leeuw and Jos Vaessen, *NONIE Guidance on Impact Evaluation*, p. 48.

[69] Jones et al., *Improving Impact Evaluation,* pp. 5 – 6.

[70] See: Howard White, *Theory-Based Impact Evaluation: Principles and Practice,* 3ie Working Paper 3 (New Delhi: 3ie, June 2009).

[71] *Ibid.* p. 18.

[72] Frans Leeuw and Jos Vaessen, *NONIE Guidance on Impact Evaluation,* p. 30.

[73] Howard White, *Theory-Based Impact Evaluation,* p. 1. TBIE is usually contrasted with the 'black box' approach, which "often simply reports an impact – being interested in the statistical significance of the coefficient for the average treatment effect, but makes no attempt to answer the why question". *Ibid.* p. 7.

[74] *Ibid.* p. 15.

[75] Drawn from: Howard White, *Theory-Based Impact Evaluation,* pp. 7 – 17. For more detail on these key steps, please refer to this working paper.

[76] *Ibid*. p. 1. For example, the NONIE guidance favours the "symbiosis between theory-based evaluation and quantitative impact evaluation." Frans Leeuw and Jos Vaessen, *NONIE Guidance on Impact Evaluation*, p. 31.

[77] John Mayne, *Contribution analysis: An approach to exploring cause and effect*, ILAC Brief, No. 16*,* (Fiumicino: Institutional Learning and Change (ILAC) Initiative c/o Bioversity International, July 2008), p.1.

[78] John Mayne, 'Addressing attribution through contribution analysis', *Discussion Paper*, Office of the Auditor general of Canada, June 1999, pp. 19 – 20.

[79] Adapted from: John Mayne, *Contribution analysis: An approach to exploring cause and effect*, ILAC Brief 16, May 2008. For a full overview of these steps please see this article.

[80] John Mayne, 'Contribution Analysis', pp. 71-72.

[81] Sarah Earl, Fred Carden and Terry Smutylo, *Outcome Mapping: Building Learning and Reflection into Development Programs* (Ottawa: International Development Research Centre, 2001), p. 1.

[82] *Ibid*. p. 1.

[83] *Ibid*. p. 1.

[84] Steve Powell and Ivona Čelebičić, *Outcome Mapping Evaluation of Six Civil Society Projects in Bosnia and Herzegovina,* Evaluation Summary Report. (Stockholm: Swedish International Development Cooperation Agency (Sida), 2008), 10.

[85] For a detailed account of the methods used see: Sarah Earl et al., *Outcome Mapping,* Appendix B, pp. 129 – 30.

[86] The three stages are adapted from: Terry Smutylo, *Outcome mapping: A method for tracking behavioural changes in development programs*, ILAC Brief 7, (Fiumicino: ILAC c/o Bioversity International, August 2005), p. 2.

[87] *Ibid.* p. 16. This was the case with the OM evaluation conducted by SIDA. Steve Powell and Ivona Čelebičić, *Outcome Mapping Evaluation in Bosnia.*

[88] Duncan Hiscock and Simon Rynn, 'Monitoring and Evaluation', *OECD Handbook on SSR*, p. 46.

[89] ODI, *RAPID Outcome Assessment* (London: ODI, 2010), p. 1.

[90] *Ibid*.

[91] Triangulation refers to 'the use of three or more theories, sources or types of information, or types of analysis to verify and substantiate an assessment.' OECD DAC, *Glossary of Key Terms.*

[92] *Ibid*.

[93] These stages are adapted from ODI, RAPID Outcome Assessment, p. 1 – 2.

[94] See: Rick Davies and Jess Dart, *The 'Most Significant Change' (MSC) Technique – A Guide to its Use,* 2005. http://www.mande.co.uk/docs/MSCGuide.htm.

[95] Maureen O'Flynn, *Impact Assessment: Understanding and assessing our contributions to change*, M&E Paper 7, (Oxford: INTRAC, 2010), p. 10.

[96] The steps are adapted from Davies and Dart, The MSC Technique, 10.

[97] Rick Davies and Jess Dart, *The 'Most Significant Change' (MSC) Technique,* p. 79.

[98] *Ibid*.

[99] *Ibid.* pp. 12 – 13.

[100] The symbol '/ ' refers to the methodology corresponding to the approach depending on the specific context. For instance, theory-based impact evaluation can correspond to the participatory approach to evaluations to the extent that efforts are made to support the participatory nature of the evaluation.

[101] This section provides an overview of the main results from the desk study on international actors' approaches to evaluation and how they have tried to address the need for impact measurement. It was not possible to access information for each actor discussed on their specific approaches to evaluating the impact of support for rule of law and security institutions. This is because the evaluation policies and guidelines of international actors are generic and do not usually single out specific areas of their work, but also because examples of evaluations of rule of law and security institutions components at the impact level were extremely difficult to come by. This section therefore limits itself to providing an overview of how international actors approach evaluation more broadly (beyond rule of law and security institutions issues, so as to include general evaluations of country programmes for instance).

[102] These two groups have been selected because they contain the actors within the UN system that are most likely to be working on rule of law and security institutions issues. Members of the SSR task force are: DPA, DPKO, ODA, OHCHR, OSAA, PBSO, UNICEF, UNDP, UNU Women, UNODC, UNFPA. Members of the UN RoL Coordination and Resource Group are: DPA, DPKO, OHCHR, OLA, UNDP, UNHCR, UNICEF, UN Women, UNODC. Actors that are part of these groups but not considered in the mapping are the Office of Legal Affairs (OLA) and the Office of the Special Adviser on Africa (OSAA), because they either do not have evaluation policies and guidance or we had no access to information on them.

[103] Office of Development Effectiveness, *Impact Evaluation*, Evaluation Best Practice Brief, (Canberra: AusAID, Australian Government, November 2006), p. 3.

[104] JICA, *Annual Evaluation Report 2010,* (Tokyo: JICA, 2010), p. 11.

[105] USAID, *Evaluation Policy, Ibid,* (Washington DC: USAID, 2011), p. 2.

[106] *Ibid.* p. 8.

[107] These techniques are: 'difference in differences', propensity score matching, regression discontinuity design and natural experiment. JICA, *Annual Evaluation Report 2010,* p. 56.

[108] *Ibid.* p. 56.

[109] United States Agency for International Development (USAID), *A Meta Evaluation of Foreign Assistance Evaluations,* Office of the Director of U.S. Foreign Assistance (Washington DC: US Department of State, June 2011), p. 37.

[110] Kim Forss, Evert Vedung, Stein Erik Kruse, Agnes Mwaiselage and Anna Nisdotter, 'Are Sida Evaluations Good Enough? An Assessment of 34 Evaluation Reports', *Sida Studies in Evaluation,* 2008:1, p. 49.

[111] *Ibid.* p. 54.

[112] *Ibid.* p. 72.

[113] Independent commission for Aid Impact, 'ICAI's Approach to Effectiveness and Value for Money', Report 1, (London: ICAI, November 2011), p. 14.

[114] *Ibid.* p. 15.

[115] Roger Drew, *Synthesis Study of DFID's Strategic Evaluations 2005 – 2010: A report produced for the Independent Commission for Aid Impact* (London: ICAI, 2011), p. 59.

[116] Mark Pearson, *Impact Evaluation of the Sector Wide Approach (SWAp), Malawi*, Final Report (London: DFID, 22 June 2010), p. 16.

[117] *Canadian International Development Agency (CIDA), CIDA Evaluation Guide, (Gatineau, QC: CIDA, 2004),* p. E-13.

[118] Stefan Molund and Görun Schill, *Looking Back, Moving Forward : Sida Evaluation Manual,* 2nd Ed. (Stockholm: SIDA, 2007), p. 34.

[119] See: Kate Butcher and Shane Martin, *Independent Progress Report of PNG Australia Sexual Health Improvement Program (PASHIP) AidWorks Initiative Number ING918* (Canberra: AusAID, 28 April 2011).

[120] Peter Ellis, *Impact Evaluation,* Office of Development Effectiveness (Canberra: Australian Agency for International Development, November 2006), p. 3.

[121] CIDA, *Evaluation Guide,* p. E-13.

[122] Note that in this table, as with tables 5 and 6, efforts are made to determine to what extent actors are using methodologies relevant for measuring impact. However, not in all cases have these methodologies been used for examining impact, they are at times used as well as components of more general evaluations. The tables are not intended to be exhaustive but highlight examples of methodologies used.

[123] Ellis, *Impact Evaluation,* pp. 1-2.

[124] AusAID, 'Evaluation Reports', http://www.ausaid.gov.au/publications/pubs.cfm?Type=PubEvaluationReports.

[125] Robyn Renneberg, Yvonne Green, Memafu Kapera and Jean-Gabriel Manguy, *Papua New Guinea Media Development Initiative 2 AidWorks Initiative No: INF757*, Evaluation Report (Canberra: AusAID, 12 May 2010), p. 30.

[126] Kate Butcher and Shane Martin, *Independent Progress Report of PASHIP.*

[127] AusAID. *Impact evaluation of the Thailand-Australia HIV/AIDS Ambulatory Care Project,* Evaluation and Review Series no. 37, May 2005.

[128] AusAID, *Fiji: Annual program performance update 2006-07* (Canberra: AusAID, 2006), p. vii.

[129] AusAID, 'Papua New Guinea: Annual program performance update 2006-07' (Canberra: AusAID, 2006), p. 32.

[130] Sue Emmott, Ken Lee, Paul Weelen, *The Capacity Building Service Centre in Papua New Guinea Independent Evaluation Report* (Canberra: AusAID, December 2009), p. 26.

[131] CIDA, *Evaluation Study of the PANAFTEL Project* (Gatineau, QC: CIDA, 1999).

[132] This is referenced by an ICAI, not a DFID document. However, ICAI is the body mandated to evaluate DFID reports except at the project level. Independent Commission for Aid Impact, 'ICAI's Approach to Effectiveness and Value for Money', Report 1, (London: ICAI, November 2011), p. 8.

[133] *Ibid*.

[134] Pearson, *Impact Evaluation of SWAp Malawi,*

[135] Howard White, Nina Blöndal, Edoardo Masset and Hugh, Waddington, *Maintaining Momentum Towards the MDGs: An Impact Evaluation of Interventions to Improve*

Maternal and Child Health and Nutrition Outcomes In Bangladesh (London: DFID, September 2005).

136 DFID, 'Research for Development Project Record: Climate Change Adaptation in Africa,' http://www.dfid.gov.uk/r4d/Project/60055/Default.aspx.

137 Jeremy Clarke et al., *DFID Influencing in the Health Sector: A Preliminary Assessment of Cost Effectiveness,* DFID Working Paper, No. 33, prepared by the Overseas Development Institute (London: DFID, November 2009).

138 Japan International Cooperation Agency, *New Guidelines for Project Evaluation,* Evaluation Department (Tokyo: JICA, 2010), p. 73.

139 JICA. *Annual Evaluation Report 2010* (Tokyo: JICA, 2010), pp. 56-57.

140 Karin Attström, Anders Kragh and Vladimir Korzh, *Young People Against Drugs – The Pinsk Model in Belarus*, Sida Evaluation 2008:22 (Stockholm: SIDA, May 2008), p. 5.

141 Steve Powell and Ivona Čelebičić, *Outcome Mapping Evaluation in Bosnia.* See also: Molund and Schill, *Sida Evaluation Manual,* p. 99.

142 USAID, *USAID Evaluation Policy* (Washington DC: USAID, 2011), p. 7.

143 Jeffrey Swedburg and Steven Smith, *Mid-Term Evaluation of USAID's Counter-Extremism Programming in Africa* (Washington DC: USAID, February 1 2011), p. 1.

144 The Mitchell Group, *Impact Evaluation of the Community Revitalization Through Democratic Action Program (CRDA), Serbia Local Government Reform Program (SLGRP), and Serbia Enterprise Development Project (SEDP)* (Washington DC: USAID, June 2 2008).

145 Fiona Kotvojs and Bradley Shrimpton, 'Contribution analysis - A new approach to evaluation in international development', *Evaluation Journal of Australasia,* vol. 7, no. 1, autumn 2007, p. 27.

146 Kotvojs and Shrimpton, 'Contribution analysis', p. 29.

147 *Ibid.* p. 33.

148 *Ibid*.

149 Steve Powell and Ivona Čelebičić, *Outcome Mapping Evaluation in Bosnia.*

150 NONIE's guidance is not intended to represent the agreed positions of all the individual NONIE members; however, the findings draw on consultation among a reference group of evaluation experts. NONIE, 'About NONIE', http://www.worldbank.org/ieg/nonie/about.html.

151 Frans Leeuw and Jos Vaessen, *NONIE Guidance on Impact Evaluation*, p. 77. The other methodologies mentioned in the annex are success case method and the method for impact assessment of projects and programmes (MAPP). They are not covered in the above discussion because they have mostly been associated with very specific contexts such as training evaluations and rural appraisal evaluations.

152 Franck Perrault, 'From the Director's Desk', in *OPEV Sharing [newsletter]*, Operations Evaluation Department, fall 2011.

153 The African Development Bank (AfDB) is a member of the Evaluation Cooperation Group (ECG) which is a member of NONIE (see below).

154 AfDB, *Uganda Poverty Alleviation Project: Project Performance Evaluation Report (PPER)*, Operations Evaluation Department (Tunis: AfDB, 14 June 2001), p. 5.

[155] AfDB, *Egypt, Morocco and Tunisia: Energy Sector Impact Study*, Operations Evaluation Department, (Tunis: AfDB, 5 June 1998).

[156] African Development Fund, *Nigeria: Rural Water Supply and Sanitation Sub-Programmes In Yobe and Osun States - Appraisal Report*, Water and Sanitation Department, (Tunis: AfDB, May 2007), pp. 37-40.

[157] See: AfDB, *Evaluation of the Decentralisation Strategy and Process in the African Development Bank: Concept Note*, Operations Evaluation Department, (Tunis: AfDB, 22 October 2008).

[158] EuropeAid, 'Evaluation of projects, programmes and strategies', http://ec.europa.eu/europeaid/evaluation/ methodology/methods/mth_pps_en.htm.

[159] EuropeAid, 'Cause-and-effect analysis', http://ec.europa.eu/europeaid/evaluation/ methodology/methods/ mth_att_en.htm.

[160] Robert N. LeBlanc et al., *Country Level Report: Uganda,* Final Report prepared for the European Commission, November 2009.

[161] Consortium composed by ECO Consult, AGEG, APRI, Euronet, IRAM, NCG, '*Country Level Report: Botswana*', European Commission, December 2009. Available at: http://ec.europa.eu/europeaid/how/evaluation/evaluation_reports/reports/2009/1273 _vol2_en.pdf.; European Commission: '*Country Evaluation Jordan*', Final Report, August 2007. Jürgen Buchholz et al. *Country Level Report: Botswana*, Final Report prepared for Europe Aid and European Commission, December 2009. http://ec.europa.eu/europeaid/ how/evaluation/evaluation_reports/reports/2007/1087_vol3_en.pdf

[162] See, for example, use of an "Expected Impact Diagram". Aide à la Décision Economique Belgium, *Evaluation of European Commission's Support to Egypt Country level Evaluation*, Final Report prepared for the European Commission, December 2010.

[163] For an example of OECD direct involvement in an outcome mapping evaluation see OECD DAC, 'Evaluating Development Impacts - An Overview*,* http://www.oecd.org/ document/17/0,3746,en_ 21571361_34047972_46179985_1_1_1_1,00.html.

[164] Swiss Agency for Development and Cooperation (SADC) and OECD, *Outcome Mapping and empowerment: the experience of SAHA in Madagascar* (Paris: OECD DAC, 2010).

[165] See: World Bank, 'Evaluation Approach', http://web.worldbank.org/external/ default/main?theSitePK=1324361 &piPK=64252979&pagePK=64253958&menuPK=5039271&contentMDK=20789685.

[166] Independent Evaluation Group (IEG), *An Impact Evaluation of India's Second and Third Andhra Pradesh Irrigation Projects* (Washington DC: World Bank, 2008).

[167] IEG, *Do Conditional Cash Transfers Lead to Medium-Term Impacts? Evidence from a Female School Stipend Program in Pakistan* (Washington DC: World Bank, 2011).

[168] IEG, *Assessing the Long-Term Effects of Conditional Cash Transfers on Human Capital: Evidence from Colombia* (Washington DC: World Bank, January 10 2011).

[169] AfDB, *Revised Guidelines on Project Completion Report (PCR) Evaluation Note and Project Performance Evaluation (PPER)*, Operations Evaluation Department (Tunis: AFDB, January 2001), pp. 64 - 65.

[170] *Ibid.* pp. 65-6.

171 AfDB, *Uganda Poverty Alleviation Project*.
172 AfDB, *Evaluation Concept Note*.
173 EuropeAid, 'Analysis'*,* http://ec.europa.eu/europeaid/evaluation/methodology/ methods/mth_ana_en.htm#01.
174 EuropeAid, 'Analysis'*.*
175 OECD DAC, *The DAC Network in Development Evaluation – Illuminating Development Results and Challenges,* Factsheet (Paris: OECD, 2010).
176 OECD DAC, *Glossary of Key Terms.*
177 The network is one of the five NONIE members. See: OECD DAC, 'Evaluating Development Impacts - An Overview', http://www.oecd.org/document/17/ 0,3746,en_21571361_34047972_46179985_1_1_1_1,00.html.
178 OECD DAC, *Guidance on Evaluating Conflict prevention and Peacebuilding Activities,* Annex 7.
179 *Ibid.* p. 10.
180 *Ibid.* p. 41.
181 See for example: Michael Bamberger, *Conducting Quality Impact Evaluations Under Budget, Time and Data Constraints* (Washington DC: World Bank, 2006).
182 On the experience of the IEG in conducting impact evaluation, see for example: Howard White, *Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank* (Washington DC: World Bank, 2006). The World Bank has developed a handbook for quantitative impact evaluation methods. See: Shahidur R. Khandker, Gayatri B. Koolwal and Hussain A. Samad, *Handbook on impact evaluation: quantitative methods and practices* (Washington DC: World Bank, 2010).
183 Howard White, *Impact Evaluation: The Experience of the IEG,* p. 20.
184 The Joint Inspection Unit (JIU) is responsible to the General Assembly of the United Nations and similarly to the competent legislative organs of those specialized agencies and other international organizations within the United Nations system that have accepted its statute. See: JIU, 'Welcome to the JIU', http://www.unjiu.org/ . Office of Internal Oversight Services (OIOS), 'Inspection and Evaluation Division', http://www.un.org/depts/oios /pages/ied.html.
185 Office of Internal Oversight Services (OIOS), *Evaluation Manual- Guidelines for the conduct of inspections and evaluations in the United Nations Office of Internal Oversight Services,* Inspection and Evaluation Division (New York: United Nations, 2009).
186 OIOS, *Evaluation Manual,* p. 49.
187 OIOS, *Evaluation Manual,* p. 51.
188 United Nations General Assembly, *Report of the Committee for Programme and Coordination*, Supplement No. 16, A/64/16, 2 July 2009, para. 38.
189 *Ibid.* para. 45.
190 *Ibid.* para. 44.
191 DPKO and Department of Field Support, *Ten-year Impact Study on Implementation of UN Security Council Resolution 1325 (2000) on Women, Peace and Security in Peacekeeping*, Final Report (New York: United Nations, 2010), p. 13.

[192] Equitas and Office of the United Nations High Commissioner for Human Rights (OHCHR), *Evaluating Human Rights Training Activities - A Handbook for Human Rights Educators* (Montreal, QC: Equitas & OHCHR, 2011), p. 110.

[193] Peacebuilding Support Office, *Peacebuilding Review* (New York: United Nations, November 2011), p. 13.

[194] United Nations Development Programme (UNDP), *Handbook on Planning, Monitoring and Evaluating for Development Results* (New York: UNDP, 2009), p. 136.

[195] UNDP, *The Evaluation Policy of UNDP*, DP/2011/3, 15 November 2010.

[196] Stuart Black, *Impact Evaluation of the Regional Corporate Social Responsibility (CSR) Project*, Final Evaluation (European Union and UNDP, September 2008).

[197] Tony Beck, Francoise Coupal, Scott Green and Christina Bierring, *Strengthening Evaluation for Improved Programming - UNFPA Evaluation Quality Assessment* (New York: United Nations Population Fund (UNFPA), December 2005), p. xi.

[198] United Nations High Commission for Refugees (UNHCR), *UNHCR's Evaluation Policy* (Geneva: UNHCR, 2010), pp. 4 – 6; UNHCR, *Measure for measure : a field-based snapshot of the implementation of results based management in UNHCR,* Initial management response, December 2010, p. 1.

[199] Paul O'Hagan, *An Independent Impact Evaluation of UNHCR's Community Based Reintegration Programme in Southern Sudan* (Geneva: UNHCR, March 2011).

[200] United Nations Children's Fund (UNICEF), *Advocacy Toolkit: A guide to influencing decisions that improve children's lives* (New York: UNICEF, 2010), pp. 62, 75.

[201] UNICEF, 'Evaluation and lessons learned', http://www.unicef.org/evaluation/index_13486.html.

[202] UNICEF, 'Communication for Development'*,* http://www.unicef.org/cbsc/index_42377.html.

[203] UNICEF and the Ministry of Foreign Affairs of the Netherlands, *Impact evaluation of drinking water supply and sanitation interventions in rural Mozambique: More than Water*, Mid-term Report (The Hague: Ministry of Foreign Affairs of the Netherlands, October 2011).

[204] UNICEF, 'Communication for Development'.

[205] Daniel McAvoy and Michelle Legu, *Solomon Islands Earthquake qnd Tsunami Disaster: An Evaluation of UNICEF's Response in the Emergency and Initial Recovery Phases* (Fiji: UNICEF Pacific Office, April 2008), p. 3.

[206] Peter Allan and IEG, *Independent project evaluation of the Support to Crime Prevention and Criminal Justice Reform - GLOT63, Global* (Vienna: United Nations Office on Drugs and Crime (UNODC), January 2012).

[207] Peter Allan, *Mid-term evaluation of project TD/RER/H22: Establishment of the Central Asian Regional Information and Coordination Centre,* Mid-term Evaluation report (CARICC and UNODC, August 2011).

[208] UNIFEM, *Monitoring and Evaluation Framework (2010-2013),* revised March 2011; Cecilia M. Ljungman, *Meta-Evaluation – UNIFEM Evaluations 2004 – 2008,* Evaluation Report (New York: UNIFEM, 2009), p. 16.

209 UNIFEM, *M&E Framework,* p. 38.

210 Julie Lafreniere, Mary Jane Real and Ricardo Wilson-Grau, *The United Nations Trust Fund to End Violence against Women – Mapping of Grantees' Outcomes (2006 to Mid-2011)* (New York: UN Women, October 2011).

211 UNIFEM, *UN Women Safe Cities Free of Violence against Women and Girls Global Programme (2010-15): Impact Evaluation Strategy* (New York: UN Women, February 2011).

212 UNDP, *M&E Handbook*, p. 136.

213 Ana de Mendoza, *UN Women Fund for Gender Equality Monitoring and Evaluation Framework (2010-2013)* (New York: UN Women, March 2011), p. 5.

214 UNHCR, *Evaluation Policy,* p. 4.

215 Michael Bamberger and Marco Segone, *How to design and manage Equity-focused evaluations* (New York: UNICEF Evaluation Office, 2011), pp. 71-2.

216 Alternative methods include single-case analysis, longitudinal designs or interrupted time series. *Ibid.* pp. 73-4.

217 There are some examples of impact assessments on the UNICEF website, particularly in the areas of health and education, and a few in the area of mine action: See for example: Joseph-Ernest Aondofe Iyo, *School Health and Physical Education Services (SHAPES) Program: An Impact Assessment*, Evaluation Report (New York: UNICEF, 2001); Taz Khaliq, *Evaluation of UNICEF's Support to Mine Action,* commissioned by UNICEF Landmines and Small Arms Team (Shrivenham, UK: Cranfield University, June 2006).

218 The JIU evaluation of RBB in peacekeeping operations notes that "evaluation is a weak area in DPKO". Ortiz and Inomata, *Evaluation of results-based budgeting in peacekeeping operations*, p. 19. See also Sarah Dalrymple, *Survey of the United Nations Organisation's arrangements for monitoring and evaluating support to security sector reform*, (London: Saferworld, January 2009), p. 6.

219 Sarah Dalrymple, *Survey on the UN arrangements for M&E in SSR,* p. 9.

220 Cecilia M. Ljungman, *UNIFEM Meta-Evaluation,* p. 14.

221 *Ibid*.

222 Tony Beck et al., *Strengthening Evaluation for Improved Programming*, p. xi. It was not possible in the limits of the study to see whether, after adopting impact as a criterion in its new policy, assessments subsequently focussed more on impact.

223 Roger Miranda, *Evaluation in UNODC: Learning and Improving through Evaluative Thinking,* Meta-evaluation conducted by the UNODC Independent Evaluation Unit (New York: United Nations, 2008), p. 8.

224 Cecilia M. Ljungman, *UNIFEM Global Meta-Evaluation 2009*, Final Report (New York: UNIFEM, May 2010), p. 8.

225 DPKO and DFS, *Ten-year Impact Study on Implementation of UN Security Council Resolution 1325*, p. 14.

226 The working group includes the United Nations Office for the Coordination of Humanitarian Affairs (OCHA), the United Nations Children's Fund (UNICEF), the World Food Programme (WFP), the Food and Agriculture Organization of the United Nations

(FAO), the UK Department For International Development (DFID), the Active Learning Network for Accountability and Performance (ALNAP), the Emergency Capacity Building Project (ECB), DARA International, Humanitarian Accountability Partnership International (HAP-I) and the Groupe Urgence Réhabilitation Développement (Groupe URD).

227 Tony Beck, *Joint Humanitarian Impact Evaluation: options paper* (New York: United Nations Office for the Coordination of Humanitarian Affairs (OCHA), 9 December 2009), p. iv.

228 *Ibid.* p. v.

229 *Ibid.* p. 18.

230 *Ibid.* p. 19.

231 United Nations General Assembly, *Report of the Committee for Programme and Coordination*, Forty-ninth session (8 June-2 July 2009), Document No. A/64/16(SUPP) (New York: United Nations, 2009).

232 Exceptions include: Steve Powell and Ivona Čelebičić, *Outcome Mapping Evaluation in Bosnia;* AusAID, *Fiji: Annual program performance update 2006-07.*

233 For example: Cecilia M. Ljungman, *UNIFEM Meta-Evaluation,* p. 39.

234 USAID, *Meta Evlauation,* p. 34.

235 OECD DAC, *Guidance on Evaluating Conflict prevention and Peacebuilding Activities,* Annex 7.

236 Eric Mvukiyehe and Cyrus Samii, *Laying a Foundation for Peace?*, p. 9.

237 Other approaches can be used to approach "but never fully attain" the counterfactual comparison. See: *Ibid*.

238 Eric Mvukiyehe and Cyrus Samii, *Quantitative Impact Evaluation in Liberia*, February 2009, p. 22

239 A concern of a UNODC meta-evaluation was: "Do senior managers mean impact or outcome? The answer has important implications on how evaluations are conducted." See: Roger Miranda, *Evaluation in UNODC*, 34.

240 Frans Leeuw and Jos Vaessen, *NONIE Guidance on Impact Evaluation,* p. 8.

241 UNDP, *Handbook on Planning, Monitoring and Evaluating,* p. 136.

242 OECD DAC, *Guidance on Evaluating Conflict Prevention and Peacebuilding Activities,* p.27

243 Sarah Dalrymple, *Survey on the UN arrangements for M&E in SSR*, p. 34.

244 Frans Leeuw and Jos Vaessen, *NONIE Guidance on Impact Evaluation,* p. 11.*Ibid*.

245 Frans Leeuw and Jos Vaessen, *NONIE Guidance on Impact Evaluation,* p. 48.

246 As discussed above, the exception would be in the case of theory-based impact evaluation which can support both accountability and learning.

247 Robert Picciotto, 'The New Environment for Development Evaluation', *American Journal of Evaluation,* vol. 28, no. 4, December 2007, p. 517.

248 See for instance: Cedric de Coning and Paul Romita, *Monitoring and Evaluation of Peace Operations.*

249 Glenda Eoyang and Tom Berkas cited in: John Mayne, 'Contribution Analysis: Addressing Cause and Effect' in Kim Forss, Mita Marra and Robert Schwarz (eds.), *Evaluating the*

Complex: Attribution, Contribution and Beyond, (Comparative Policy Evaluation vol. 18), p. 83. (New Brunswick, NJ: Transaction Publishers, 2011).

[250] Frances Westley, Brenda Zimmerman and Michael Quinn Patton quoted in John Mayne 'Contribution Analysis', p. 83.

[251] Simon Rynn with Duncan Hiscock*, Evaluating for security and justice: Challenges and opportunities for improved monitoring and evaluation of security system reform programmes,* (London: Saferworld, December 2009), p. 17.

[252] *Ibid.* p. 16.

[253] Simon Rynn with Duncan Hiscock*, Evaluating for security and justice,* p. 18.

[254] For instance, in the case of the European Commission's evaluation of its Jordanian country programme, the approach used was to identify main projects that have been implemented and among these to select the most costly ones to assess using contribution analysis. See. *European Commission, Country Evaluation Jordan*, Volume 3, August 2007, p. 8.

[255] Sarah Earl, Fred Carden and Terry Smutylo, *Outcome Mapping,* p. 15.

[256] CDA, *Encouraging Effective Evaluation of Conflict Prevention and Peacebuilding Activities,* p. 9.

[257] USAID, *A Meta Evaluation of Foreign Assistance Evaluations,* p. 18.

[258] The table is based on analysis of the guidance material available on these methodologies, combined with examples from specific evaluations using them. Where no information was found, the field is left blank.

[259] For example, time required is noted to take from less than 6 months in the case of rapid evaluations to over two years. World Bank, *Monitoring & Evaluation: Some tools, methods and approaches* (Washington, DC: World Bank, 2004), p. 23.

[260] For example, they can cost up to 900,000 USD however, some simple evaluations can be conducted for less than 100,000 USD. *Ibid*.

[261] Centre for the Study of Violence (CSVR) and IDRC, 'Evaluating Experiences in transitional Justice and Reconciliation: Challenges and opportunities for Advancing the Field'*,* Workshop Report, April 2007, p. 7.

[262] Introduction workshop costs about US$500 per participant. See OM Workshop programme under: http://www.outcomemapping.ca/download.php?file=/resource/files/OM%20event%20in%20March.pdf.

[263] The SIDA evaluation looked at six projects and included several rounds of field visits and approx. 80–100 working days. Steve Powell and Ivona Čelebičić, *Outcome Mapping Evaluation in Bosnia,* p. 32.

[264] Rick Davies and Jess Dart, *The 'Most Significant Change' (MSC) Technique,* p. 12.

[265] Michael Bamberger, *Institutionalizing Impact Evaluation within the framework of A Monitoring and Evaluation System¸* Independent Evaluation Group (Washington DC: World Bank, 2009), p. 11.

[266] NONIE and the World Bank have been at the forefront of improving the quality of IEs and make them usable in real world conditions. See for example: Michael Bamberger,

Conducting Quality Impact Evaluations Under Budget, Time and Data Constraints, Independent Evaluation Group (Washington DC: World Bank, 2006); Michael Bamberger, Jim Rugh, Linda Mabry, *Real World Evaluations: Working under Budget, Time, Data and Political Constraints* (Thousand Oaks, CA: Sage Publications, 2006). Michael Bamberger, *Institutionalizing Impact Evaluation,* p. 12.

[267] Michael Bamberger, *Conducting Quality Impact Evaluations,* p. 7.

[268] Michael Bamberger, *Institutionalizing Impact Evaluation,* p. 12.

[269] Michael Bamberger, *Conducting Quality Impact Evaluations,* p. 6.

[270] Sarah Earl, Fred Carden and Terry Smutylo, *Outcome Mapping,* p. 15.

[271] For example, in Sida's outcome mapping evaluation, the approach was adapted to use questionnaires that examine progress markers rather than requiring boundary partners to collect the information themselves. Steve Powell and Ivona Čelebičić, *Outcome Mapping Evaluation in Bosnia.*

[272] World Bank, *Monitoring & Evaluation*, p. 6. According to the OECD DAC Glossary, "Performance measurement is a system for assessing performance of development interventions against stated goals." OECD DAC, *Glossary of Key Terms*, p. 24.

[273] UNDP, *Handbook on Planning, Monitoring and Evaluating for Development Results* (New York: UNDP, 2009), p. 65.

[274] These indicator sets, developed in 2009 cover five sectors of assistance: transport, water & sanitation, education, health and agricultural & rural development. For health sector see for example: European Commission EuropeAid, *Outcome and Impact level Intervention Logic and Indicators – Health Sector,* Working Paper, EC External Services Evaluation Unit (Brussels: European Commission External Services, October 2009).

[275] *Ibid.* p. 7.

[276] *Ibid.* p. 8.

[277] UN Inter-Agency Working Group on DDR, 'Module 3.50 Monitoring and Evaluation of DDR Programmes', in *Integrated DDR Standards (IDDRS),* 2006. Online version available at: http://unddr.org/iddrs/03/50.php.

[278] *Ibid*.

[279] *Ibid*.

[280] *Ibid*.

[281] UNDP and BCPR, *How to Guide - Monitoring and Evaluation for Disarmament, Demobilization and Reintegration Programmes* (New York: UNDP/BCPR, 2009), p. 26.

[282] CDA, *Encouraging Effective Evaluation of Conflict Prevention and Peacebuilding Activities,* p. 28.

[283] Ken Menkhaus, *Impact Assessment in Post-Conflict Peacebuilding*, p. 4.

[284] Christoph Spurk,'Forget impact', p. 13.

[285] CDA, *Encouraging Effective Evaluation of Conflict Prevention and Peacebuilding Activities,* p. 28.

[286] Kenneth Watson, *UNICEF Country Offices Evaluations,* p. 14.

[287] Cecilia M. Ljungman, *UNIFEM Global Meta-Evaluation 2009*, Final Report (New York: UNIFEM, 2010), p. 48.

[288] Anette Wenderoth et al., *UNIFEM Programme Facilitating CEDAW Implementation in Southeast Asia,* Evaluation Report (New York: UNIFEM, 2008), pp. 4, 143.

[289] World Bank, 'Law & Justice Institutions - Performance Measures Topic Brief', http://web.worldbank.org/ WBSITE/EXTERNAL/TOPICS/EXTLAWJUSTINST/0,,contentMDK:20756997~menuPK:20256 88~pagePK:210058~piPK:210062~theSitePK:1974062~isCURL:Y,00.html

[290] Outcome Mapping Online Community, 'Outcome Mapping FAQs', http://www.outcomemapping.ca/about/faqs.php .

[291] Cedric de Coning and Paul Romita, *Monitoring and Evaluation of Peace Operations*, p. 1.

[292] Aide à la Décision Économique (ADE) Belgium and PARTICIP, *Thematic Evaluation of the European Commission support to Conflict Prevention and Peace Building – Final Report,* Vol. I: Main Report, Evaluation for the European Commission, (Brussels: European Commission, October 2011), 92.

[293] *Ibid.*, 98-9.

[294] Simon Rynn with Duncan Hiscock*, Evaluating for security and justice,* p. iv.

[295] GSDRC, *M&E in Fragile States*, p. 1.

[296] CDA, *Encouraging Effective Evaluation of Conflict Prevention and Peacebuilding Activities,* p. 34.

[297] *Ibid*.

[298] Kenneth Watson, *UNICEF Country Offices Evaluations,* p. iv.

[299] *Ibid*.

[300] United Nations General Assembly, *Triennial comprehensive policy review of operational activities for development of the United Nations system,* Resolution adopted by the General Assembly, Document No. A/RES/59/250, (New York: United Nations, 17 August 2005), pp. 11-12.

[301] See: OECD DAC, *Quality Standards in development evaluation* (Paris: OECD, 2010), p. 7. These standards are based on the OECD DAC, *Paris Declaration on Aid Effectiveness: Ownership, Harmonisation, Alignment, Results and Mutual Accountability* (Paris: OECD, 2005).

[302] OECD DAC, *Quality Standards in development evaluation,* p. 7.

[303] Participatory evaluation approaches can 'empower local people to initiate, control and take corrective action'. Cecilia M. Ljungman, *UNIFEM Meta-evaluation*, p. 38.

[304] OECD DAC, *Guidance on Evaluating Conflict Prevention and Peacebuilding Activities*, Annex 7, p. 86.

[305] Simon Rynn with Duncan Hiscock*, Evaluating for security and justice,* p. 11.

[306] See: Cecilia M. Ljungman, *UNIFEM Meta-evaluation*, p. 34.

[307] Hans Lundgren and Megan Kennedy, 'Supporting Partner Country Ownership and Capacity in development Evaluation, The OECD DAC Evaluation Network' in Marc Segone (ed.), *Country-led monitoring and evaluation system,* (New York: UNICEF, 2009), p. 92.

[308] OECD DAC, *Guidance on Evaluating Conflict Prevention and Peacebuilding Activities*, Annex 7, p. 86.

[309] Frans Leeuw and Jos Vaessen, *NONIE Guidance on Impact Evaluation,* p. 48.

[310] Tony Beck, Francoise Coupal, Scott Green and Christina Bierring, *Strengthening Evaluation for Improved Programming - UNFPA Evaluation Quality Assessment* (New York: UNFPA, December 2005), p. xii.

Measuring the Impact of Peacebuilding Interventions on Rule of Law and Security Institutions

Vincenza Scherrer

Since the 1990s, internationally-supported peacebuilding interventions have become increasingly prominent. Activities focusing on rule of law and security institutions are a key component of this agenda. Despite increasing calls for more rigorous analysis of the impact of peacebuilding interventions, conceptual advances have been limited. There is little clarity on what is working, what is not, and why. This SSR Paper seeks to address this gap by mapping relevant approaches and methodologies to measuring impact. It examines how international actors have approached these questions in relation to support to rule of law and security institutions in complex peacebuilding environments. Most significantly, the paper demonstrates that measuring impact is not only feasible but necessary in order to maximise the effectiveness of major international investments in this field.

Vincenza Scherrer is Programme Manager of the United Nations and Security Sector Reform programme at the Geneva Centre for the Democratic Control of Armed Forces (DCAF). She has provided support to UN entities in the areas of policy research and guidance development on topics ranging from national security policy making to the nexus between disarmament, demobilization and reintegration and security sector reform. Prior to joining DCAF, Vincenza worked at the UNDP's Bureau for Crisis Prevention and Recovery and at the UK Parliament. Vincenza holds degrees from the Graduate Institute of International Studies in Geneva and the London School of Economics.

9 789292 222024