



Department
for International
Development



Measuring the Un-Measurable

Solutions to Measurement Challenges in
Fragile and Conflict-affected Environments



Principal Authors

Vanessa Corlazzoli and Jonathan White
Search for Common Ground

Date:
March 2013

Practice Products for the CCVRI

Improving Measurement in DFID Crime, Conflict & Violence Programming

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

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

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

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

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

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

Members of the consortium

Document Summary

Title:

Measuring the Un-Measurable: Solutions to Measurement Challenges in Fragile and Conflict-affected Environments

Purpose and intended use of this document:

The purpose of this practical how-to guide is to provide an overview of key tools, methodologies and approaches in the social sciences that can be utilised for measuring intangible change in conflict-affected and fragile environments. The paper presents the tool and examines the strengths and weaknesses of the individual tools.

Each tool is accompanied with a small discussion on how the tool can be used for measuring hard-to-reach, intangible changes in situations of conflict and fragility. When appropriate, examples on how to apply the tools in conflict and fragile environments have been integrated.

Key questions this document addresses:

- Why are programme results difficult to measure in conflict, crime and security?
- What are the existing social science tools and methodologies that can help us measure results in conflict, crime and security?
- How can these tools be best used, and for what?

Key messages/essential “take aways”:

- The primary focus of this paper is how to measure what is often considered un-measurable.
- The challenge is how do we collect information for intangible changes in attitude and behaviour in complex conflict and fragile environments?
- No single tool, method or approach is a cure-all panacea to the measurement challenges in peace & conflict and security & justice. Instead one will need to mix tools, methods, and approaches.
- Each tool has its own strength and weakness, and it is only by identifying what one needs to know that it becomes easier to pair methodologies.
- Any intervention occurring in a fragile or conflict-affected environment must include a robust conflict and context monitoring, a capacity building strategy for staff, and a monitoring and evaluation plan. The importance of this cannot be understated and is fundamental to interventions occurring in such environments.

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

The primary audience of this document are DFID advisers and Monitoring and Evaluation Technical staff designing or implementing programming or strategies in fragile and conflict affected states (FCAS). The secondary audience are implementing partners in FCAS. This document is probably not useful for individuals with a high level of technical expertise in monitoring and evaluation.

The document assumes some basic knowledge of design, monitoring and evaluation. It can be read either from start to finish, or as independent sections. It assumes that the reader has an intermediate knowledge of social science research methodologies and tools. It also assumes that the reader has an

intermediate knowledge of the challenges of intangible measurement in situations of conflict and fragility.

Key topics/tags:

Conflict and fragility

Evaluation

Measurement

Tools

Methodology

Social Science

Outcome

Impact

Peace Writ Large

Authors and their organizations:

Vanessa Corlazzoli, Search for Common Ground

Jonathan White, Search for Common Ground

We are grateful to Search for Common Ground's Research Assistant, Valerie Oliphant, as well as expert reviewers: Jennifer Leith (DFID), Rachel Goldwyn (CARE International UK), Kelsi Stine (SFCG), and Issaka Traore (SFCG).

Cross-references to other documents in the series:

Bayne, Sarah. *M&E of Integrated Peacebuilding and Development*. DFID: Department for International Development, 2013.

Corlazzoli, Vanessa, and Jonathon White. *Back to Basics: A Compilation of Best Practices in Design, Monitoring & Evaluation in Fragile and Conflict-affected Environments*. DFID: Department for International Development, 2013.

Corlazzoli, Vanessa, and Jonathon White. *Practical Approaches to Theories of Change in Conflict, Security & Justice Programmes Part II: Using Theories of Change in Monitoring and Evaluation*. 2012.

Goldwyn, Rachel. *Making the Case for Conflict Sensitivity in Security and Justice Sector Reform Programming*. DFID: Department for International Development, 2013.

Goldwyn, Rachel and Diana Chigas, *Monitoring and Evaluating Conflict Sensitivity: Methodological Challenges and Practical Solutions*. DFID: Department for International Development, 2013.

Woodrow, Peter. *Practical Approaches to Theories of Change in Conflict, Security & Justice Programmes Part I*. 2012.

Table of Contents

Solutions to Measurement Challenges in Fragile and Conflict-affected Environments .. 1

Introduction..... 6

Structure of the Paper 7

Challenges to Measuring Impact in Peace & Conflict, Security & Justice 9

What is Peace and Conflict? What to measure? 10

Conflict Drivers and Factors 12

Determining the Complexity of the Problem and Programme 12

Social Science Solutions 14

PART I: SOCIAL SCIENCE METHODOLOGY FOR OUTCOME MEASUREMENTS..... 15

Likert scales..... 15

Community Score Cards..... 17

Indices..... 19

Proxy Indicators 20

Information and Communications Technologies..... 22

Incident Log..... 24

Media Content Analysis and Discourse Analysis..... 24

Participant Diaries and Video Logs..... 26

Rapid Assessment Procedures..... 28

Remote Monitoring 29

Social Network Analysis..... 30

Stakeholder Analysis 31

Outcome Mapping 32

PART II: SOCIAL SCIENCE METHODOLOGY FOR IMPACT-LEVEL AND LONG TERM

MEASUREMENT..... 35

Impact Evaluation, Quasi- and Experimental Methods 35

Natural Experiments..... 36

List Experiments..... 37

Most Significant Change 39

Global Giving Storytelling Methodology 39

Systems analysis..... 41

PART III: SOCIAL SCIENCE METHODOLOGY FOR MEASURING CONTRIBUTION TO PEACE

WRIT LARGE 43

Longitudinal and Cohort Studies..... 43

Meta-Analysis 44

Conclusion 45

Annex A: List of Relevant Indices..... 46

Bibliography 49

Introduction

The international community has devoted increasing resources to address deep-seated societal antagonism in situations of conflict and fragility. Approximately 38% of all overseas direct assistance—or \$50 billion—is devoted to fragile and conflict-affected states.¹ Yet not one of these states has achieved a single Millennium Development Goal.² With the international community currently re-envisioning the next iteration of international assistance targets, including the prospect of direct targets for democracy, the security sector, governance, and justice, the need for rigorous evidence of ‘what works, what doesn’t and under what conditions’ is all the more essential.

To improve overall societal stability and resilience in situations of conflict and fragility, interventions will need to include a range of interconnected sectors, including peacebuilding, justice, security, and crime.³ In order to make a contribution towards development, these programmes will have to be able to monitor their contributions, measure their impacts, and learn from their results.

While nuances exist between the types of interventions, accurate and rigorous measurement of desired change is a common challenge for any sector working in conflict and fragile environments. The changes sought in knowledge, attitude, behaviour and perceptions are not easily grasped, observed or measured; they are inherently intangible. This is compounded by obstacles that prevent or hinder effective and rigorous data collection techniques and evaluation methods in conflict and fragile situations (FCAS).

Of course, many measurement challenges can be mitigated, and even avoided, with thorough and intentional programme design. Principles of good programme design have been long established⁴; so too have the common issues and shortfalls of programme designs in peace & conflict⁵ and security & justice.⁶ Some of these challenges include: implicit or unclear theories of change, overly ambitious and often unachievable goals and objectives, poor indicators, emphasis on output rather than outcome indicators, and poor, implicit or missing conflict analyses. Design obstacles are beyond the scope of this paper and best practices have been written and compiled into manuals,⁷ how-to notes,⁸ and publications. Instead, this paper assumes ‘good enough’ design principles have already been applied to a programme.

Measurements of results in conflict and fragile situations can be improved by increasing the research capacity of local staff, civil society organizations, research institutes, and statistical ministries or

¹ *Fragile states 2013: Resource flows and trends in a shifting world*. OECD-DAC International Network on Conflict and Fragility (INCAF), 2012.

² *Fragile States 2013*, OECD-DAC.

³ What is referred to henceforth in this paper as peace & conflict, security & justice.

⁴ Church, Cheyanne, and Mark Rogers. *Designing for Results: Integrating Monitoring and Evaluation in Conflict Transformation Programs*. Washington, DC: Search for Common Ground, 2006; *Evaluating peacebuilding activities in settings of conflict and fragility improving learning for results*. Paris: OECD-DAC, 2012; *Interim Guidance Note: Measuring and managing for results in fragile and conflict-affected states and situations*. London: DFID: Department for International Aid, 2010; *Reflecting on Peace Practice: Participant Training Manual*. Cambridge: CDA Collaborative Learning Projects, 2009.

⁵ See, for example, *Evaluating peacebuilding activities*, OECD; *Reflecting on Peace Practice*, CDA Collaborative Learning Projects; Church, Cheyanne A. "Peacebuilding Evaluation: From Infancy to Teenager." *New Routes* 13, 2008: 3-6; Scharbatke-Church, Cheyanne. "Evaluating Peacebuilding: Not Yet All It Could Be." *Berghof Handbook for conflict Transformation, Section II: Analysing Conflict and Assessing Conflict Transformation*, Berghof Foundation: 459-482; Corlazzoli and White, *Back to Basics*.

⁶ See, for example, *Governance, Security and Justice: Program Prospectus for 2011-2016*. DC: International Development Research Centre, 2011.; Stone, Christopher. *Supporting Security, Justice, and Development: Lessons from a New Era*. New York: Vera Institute of Justice, 2005.

⁷ Corlazzoli and White, *Back to Basics*.

⁸ DFID How To Note on Measuring and Using Results in Security and Justice Programmes; DFID How To Note on M&E of Employment Promotion Interventions in FCAS; DFID How to Note on the Production of Information for Policy Making in Countries Emerging from Conflict; DFID How to Note: A Practical Guide to Assessing and Monitoring Human Rights in Country Programmes; DFID How To Note on Measuring Influencing.

departments. This includes improving educational and professional trainings, coaching and mentoring. For example, UNICEF has implemented a systems approach to developing context-specific evaluation capacity at the individual and institutional level, as well as seeking to create an environment conducive to monitoring and evaluation. This initiative was begun in early 2010, and there is still work to be done to move them from their minimum requirement of “not undermining country-led evaluation systems.”⁹ For instance, while their conceptual framework is complemented by an interactive web platform that offers tools, guides, and handbooks on M&E, it is not clear whether these web resources are easily acceptable in the current technological environment of most developing countries, where internet access may be sporadic or limited. Moreover, each UN entity places varying degrees of emphasis and priority on M&E.¹⁰

Literature abounds on best practices for the staple social science tools, methods and approaches. These include focus groups, interviews, surveys, case studies, and observation. These tools should be rigorously utilised for appropriate research questions. National researchers and monitoring and evaluation staff should increase their understanding of strengths and weakness of such tools, as well as execute the planning, drafting, analysing and reporting of findings.

The primary focus of this paper is instead on how to measure what is often considered ‘un-measurable’: the intersection of conducting measurement for intangible changes in attitude and behaviour in complex conflict and fragile environments. The paper additionally explores the use of social science methodologies to understand how these intangible changes scale up to systematic/system-wide change in complex systems.

Structure of the Paper

This paper builds on other publications in DFID’s *Results Initiative in Conflict, Crime, and Violence Programming* by examining the key tools for measuring intangible change in peace, conflict, security & justice programming.

First, the paper outlines key challenges to measurement, including an examination of the key theories underpinning programming and inherent challenges to evaluation and attribution. It introduces a conceptual framework for breaking down evaluation challenges into more manageable pieces (simple, complicated, complex), which is applied throughout the remainder of the paper.

Second, and to the primary purpose, the paper examines a range of measurement tools and approaches from across the social sciences that can be adapted to evaluating intangible change. This second section is broken into three parts: tools for outcome measurement; tools for measuring impact and change over time; and tools for measuring contribution to peace writ large.

Many of the tools can be adapted for various monitoring and evaluation approaches and methodologies, and therefore their placement in one category of measurement should not be viewed as static. The tools outlined can and have been successfully applied in settings of conflict and fragility and examples are included in this paper. DFID may consider tools at the impact and peace writ large levels in

⁹ United Nations Children's Fund. "Evidence for Children: Developing national capacities for Country-led evaluation systems." United Nations Children's Fund Evaluation Office. www.unicef.org/evaluation/files/DevelopingCapacities_CountryEvaluationSystems_final.pdf (accessed March 20, 2013).

¹⁰ See UNEG publication for overview of evaluation units in 43 UN entities: United Nations Evaluation Group. "Evaluation Capacity in the UN System." United Nations Evaluation Group. www.unicef.org/evaluation/files/Evaluation_Capacity_in_the_UN_System.pdf (accessed March 20, 2013).

particular for its increasing use and potential standardisation of ex-post evaluations.¹¹

It should be noted, however, that the tools presented in this paper are uniquely suited for particular measurement challenges and evaluation questions. They cannot simply be applied 'out of the box'. Each tool must be researched, altered, and adapted to fit the unique circumstance in which it will be applied so that it takes into account gender-dynamics, conflict sensitivity, and cultural appropriateness.

Finally, successful measurement of the un-measurable does not rely on using one single tool but identifying and effectively implementing a complimentary suite of tools that in combination will lead to the un-measurable.

¹¹ *Interim Guidance Note: Measuring and managing for results*, DFID, 27.

Challenges to Measuring Impact in Peace & Conflict, Security & Justice

Numerous challenges face the measurement of results of peace and conflict, security, and justice programming.

First, the *environments* in which these interventions occur are complex. The programmes may even be adaptive and conflicts emergent. Data collection is constrained both by access to individuals, communities, regions, and propriety of culture and context, including trauma. Many, if not most, of these environmental challenges can be overcome if resources are abundant and staff have the necessary skills. In conjunction with other challenges, however, the task is far less simple.

Second, the theories of change that underpin peace & conflict and security & justice programmes seek results that are rarely tangible, countable, or knowable by the beneficiaries or population at large. The desired results of these types of programmes are intangibles, such as shifts in trust, tolerance, justice, empowerment, or acknowledgement between individuals and groups may have been adversaries in recent or historical memory. These programmes seek a change in knowledge, attitude, behaviour, and perceptions of 'the other' or a complex issues in a short period of time. They may also seek to cause ripple effects in complicated causal-relationships and ultimately transform cultures, norms, and environments to decrease the likelihoods of violent conflict. However, these changes rarely manifest within the immediate scope of the programme.

Third, intangible change is compounded with the complexity of the causal mechanisms that bring about change. It is difficult to measure how changes in values and perceptions manifest themselves, (directly or indirectly, intended or unintended) as behaviour change at the individual-, group- and/or societal-levels. How these changes interact with other elements of societal systems, for instance institutional reform or electoral outcomes, is even harder to measure. For example, an intervention might (1) create a radio programme in order to promote attitude shifts towards reconciliation among ethnic groups, while (2) simultaneously work through youth councils to advance peaceful democratic reform. In the evaluation of this intervention, it would be difficult to determine whether either or both of these activities led to a change in attitude that in turn led to a general decrease in violence in the next election. Furthermore, most programmes are based on experimental designs that have not been well vetted, tested and evaluated. Lessons learned from one context may not be applicable to another conflict dynamic. As can be seen, in the midst of intangible change and complex environments, experimental programme designs, and causal mechanisms, the measurement and attribution of peace & conflict and security & justice programmes is exceedingly difficult.

Fourth, programmes in these sectors are frequently ill-defined. In some cases programmes are designed to be adaptive and emergent in order to reflect the environment in which it occurs. Allowing for flexibility, intentionally adaptive designs facilitate quick response times from the programme as new opportunities and dynamics arise. The adaptability of these interventions may very well be the key to their success, and as a result not all programmes have pre-defined outcomes or causal pathways. This requires additional efforts to elicit evaluative information from programme documents, staff and beneficiaries to measure results.

More often than not, however, peacebuilding and statebuilding programmes are unintentionally poorly defined: unclear or overly-broad purpose statements; implicit, unarticulated and/or untested theories of change; lack of explicit conflict analyses; poor budgeting for M&E processes; lack of gender-analysis or gender mainstreaming in activities, among others. Regardless of intentionality, ill-defined

programming poses a major challenge to evaluation processes and makes attribution of any form extremely difficult to measure.

Finally, an on-going debate in the peace & conflict and justice & security fields has resulted in a shift in mind-set from measuring 'attribution' to 'contribution.'¹² *Attribution* is when it is possible to "demonstrate a direct causal link between an intervention and its impact."¹³ When working in conflict and fragile states, many things are often happening at the same time-- multiple programs and organizations may be working in the same area, the context and conflict environments may rapidly evolve or devolve, programmes may have more than one causal strand. This makes it difficult to prove that your specific intervention was the one that brought about peace writ-large. For this reason, practitioners and donors are beginning to recognize that it is better to monitor and evaluate how their interventions and actions *contribute* to an increase in peace or a decrease in violence. In other words, donors are moving toward recognizing that their actions are one of many contributing factors effecting peace. This shift from attribution to contribution is also more empowering for local beneficiaries and stakeholders, including national governments, who will not feel that international interveners are taking all of the credit for development in their country. Recognising the need to distinguish between the two, this paper will be focusing on how to measure *contribution* to peace.

Cumulatively, these four types of challenges pose particular difficulty to the measurement and evaluation of results of peace & conflict, security & justice programming in fragile and conflict-affected environments.

What is Peace and Conflict? What to measure?

Peace, broadly defined by Johan Galtung, is "a general expression of human desires, of that which is good, that which is ultimately to be pursued."¹⁴

Breaking this concept down reveals two categories of peace:

Negative peace is the *absence* of violence, including interpersonal, intergroup, and interstate violence.¹⁵

- Relatively easy to measure: homicide rates, frequency and intensity of intergroup violence, with some nuances such as perceptions of intergroup tensions and grievances.

Positive peace is the *presence* of harmonious and positive human relations at all levels¹⁶

- Harder to measure due to lack of conceptual clarity¹⁷ and the intangible nature of sub-components, such as "presence of cooperation, freedom from fear, freedom from want, economic growth and development... equality, justice, freedom of action, pluralism, dynamism" and the "absence of exploitation."¹⁸

¹² deConing, Cedric and Paul Romita. *Monitoring and Evaluation of Peace Operations*. New York: International Peace Institute, 2009: 4.

¹³ Scherrer, Vincenza. *Measuring the Impact of Peacebuilding Interventions on Rule of Law and Security Institutions*. Geneva: Geneva Centre for the Democratic Control of Armed Forces, 2012: 8.

¹⁴ Galtung, Johan. *Theories of Peace: A Synthetic Approach to Peace Thinking*. Oslo: International Peace Research Institute, Oslo, 1967: 6.

¹⁵ Galtung, *Theories of Peace*, 12.

¹⁶ Galtung, *Theories of Peace*, 12.

¹⁷ Abdi, Dekha Ibrahim. "Peace, Peacebuilding, Peacemaking." In *Berghof handbook for conflict transformation*. Berlin: Berghof Research Center for Constructive Conflict Management, 2001.

¹⁸ Galtung, *Theories of Peace*, 14-15 for further definitions and exploration of these concepts.

The absence of violence or a decrease in negative peace does not mean that there is positive peace. For example, direct, inter-group violence may cease, but positive peace will still be elusive without reconciliation and redress of key grievances, which led to and fuelled the violence in the first place. Measures of positive peace are required, many of which are intangible.

It is noteworthy that not all of the characteristics of positive peace are intangible (economic development, for example). This provides opportunities for data triangulation, the use of proxy indicators, and ripple effects throughout systems for unintended and indirect consequences.

Positive peace can be defined as the absence of the three types of violence: (1) direct/interpersonal, (2) structural, and (3) cultural:

Direct violence, or “assaults on bodily integrity” including the denial of physical resources needed for survival¹⁹, is relatively easy to measure through its tangible nature. For instance, it is relatively easy to measure the number of children that were murdered over the past three months.

Structural violence is more difficult to measure, as it “refers to conditions that both jeopardize bodily integrity... and deny other basic human needs for identity and security.” In other words, structural violence looks at how institutions and structures have impaired the ability of an individual to grow and prosper. This includes institutionalised aspects of society such as racism, sexism, ethnocentrism, a caste system, as well as inequitable laws related to access to education or healthcare. Structural violence can be difficult to measure because it is hard to prove what would have happened had those impediments not existed. Institutions are also part of a larger system, making them hard to isolate as a cause.

Cultural violence is harder to measure: it is the “aspects of our meaning [generating] systems – especially those gathered in religion, political ideologies, science, art, and media more generally – that *legitimize direct and structural violence* and perpetuate militarism.”²⁰ As such, much of cultural violence lays in values, perceptions, and worldviews which advantage one group of people over another. For instance, a society may be indifferent to the deaths of some children, given their ethnic or religious groups. This society may discriminate against certain groups and justify the deaths. The intangible nature of values and culture makes these concepts harder, but not impossible, to measure.²¹

Practitioners increasingly create multidisciplinary programmes with multiple approaches to direct, structural, and cultural violence. Some aspects of the programmes and their outcomes and impacts may be easier to measure than others.

Most disciplines still struggle to measure their impact on Peace Writ Large.²² Such sectors include security, justice, human rights, reconciliation, and other umbrella concepts, such as human security.²³

¹⁹ Barnes, Catherine. *Theories of Conflict: Causes, Dynamics and Implications, Applied Conflict Transformation Studies Series*. Birmingham UK: Responding to Conflict, 2005: 29, summarizing John Galtung’s various works referring to direct violence

²⁰ Barnes, *Theories of Conflict*, 29; italics added.

²¹ Miall, Hugh, Oliver Ramsbotham, and Tom Woodhouse. *Contemporary conflict resolution: the prevention, management and transformation of deadly conflicts*. Cambridge, UK: Polity Press, 1999: 15.

²² Defined as societal-level outcomes and impacts relating to peace beyond the micro-level of a single project, see *Reflecting on Peace Practice*, CDA Collaborative Learning Projects, 2.

²³ See Alkire, Sabina. *A Conceptual Framework for Human Security: Working Paper 2*. Oxford: Centre for Research on Inequality, Human Security and Ethnicity, University of Oxford and the Department for International Development, 2003.

When taken together at a macro-level, this can be seen as a systems-based approach²⁴ to societal transformation. Take, for example, the international community's response to the earthquake in Haiti in 2010. Donors funded a range of activities, from emergency and humanitarian response, to conflict management, health, education, and governance. The short term goal was reconstruction of Haiti, but the ultimate goal which all the various streams of activities worked towards was a more stable and resilient society that could withstand systemic shocks, i.e., Peace Writ Large.

Considerations of unintended and indirect consequences of outcomes and impacts become even more essential with such integrated multi-sectoral approaches and further compounds measurement challenges facing fields that seek impact on the umbrella concept of peace.

Conflict Drivers and Factors

Any intervention occurring in a fragile or conflict-affected environment must include a robust conflict and context monitoring and evaluation plan. *The importance of this cannot be understated and is fundamental to interventions occurring in such environments.*²⁵

An intervention taking place in contexts of conflict and fragility should develop key indicators to track the evolution of conflict and context dynamics. These indicators can help identify potential turning points for the context and its implications for the intervention. Sudden decisions to re-orient programming from one type of activity or one region to another must be based on evidence—of needs and changing and emerging dynamics.²⁶

Similarly, it is essential to have documented monitoring data tracing the evolution of conflict and context dynamics alongside the evolution of the intervention. Such information will help the evaluator make determinations on the OECD-DAC evaluative criteria for development co-operation.²⁷

Determining the Complexity of the Problem and Programme

Numerous scholar-practitioners have suggested ways to overcome the evaluation challenges listed above. Patricia Rogers,²⁸ Patricia Rogers and Richard Hummelbrunner²⁹, and Sholom Glouberman and Brenda Zimmerman³⁰ suggest distinguishing between that which is simple, complicated and complex in programme theory.

Simple problems are when “the destination is known, the path to reach it is known, and it is simple to follow the plan.”³¹ Simple problems are predictable and the outcome is known. For example, following a

²⁴ See, for example, Diamond, Louise, and John W. McDonald. *Multi-track diplomacy: a systems approach to peace*. 3rd ed. West Hartford: Kumarian Press, 1996.

²⁵ *Building Peaceful States and Societies*, DFID; *Working Effectively in Conflict-affected and Fragile Situations Briefing Paper I: Monitoring and Evaluation: A DFID Practice Paper*. London: DFID: Department for International Development, 2010; *Fragile states 2013*, OECD-DAC.

²⁶ For more on how to do this, see: Goldwyn and Chigas, *Monitoring and Evaluating Conflict Sensitivity*; Picard, Mary, and Middle East. "Measurement and methodological challenges to CARE International's rights-based programming." *Enterprise Impact News* 33, 2004; Corlazzoli and White, *Back to Basics*; Church and Rogers, *Designing for Results*.

²⁷ *Evaluating Development Co-Operating Summary of Key Norms and Standards*. 2nd ed. OECD-DAC Network on Development Evaluation, 2010.

²⁸ Rogers, Patricia J. "Using programme theory to evaluate complicated and complex aspects of interventions." *Evaluation* 14, no. 1, 2008: 29-48.

²⁹ Rogers, Patricia, and Richard Hummelbrunner. "On-line e-Learning programme on: Module 1 Equity-focused Evaluations."

³⁰ Glouberman, Sholom, and Brenda Zimmerman. "Complicated and complex systems: what would successful reform of Medicare look like?." *Changing Health Care in Canada: The Romanow Papers* 2, 2002: 21-53.

³¹ Rogers and Hummelbrunner, *On-line e-Learning programme*, 124.

recipe to baking a cake is a simple problem. *It is what you can measure.* An example could be knowledge change resulting from a well-tested and delivered training.

Complicated problems refer to interventions in which multiple components are required to produce the intended result (multi-step causal chains). For example, a complicated problem might be an evaluation of a campaign to create favourable public opinion resulting in the passing of an anti-discrimination law. The intervention involved multiple elements, all of which need to mesh together in just the right way in order to arrive at the desired result.³² Such problems may require additional measurement effort,³³ such as the use of multiple tools to triangulate data to arrive at conclusions. Such triangulation might be applied to multiple causal chains that occur simultaneously, which suggests that the programme employed more than one overall theory of change. Alternatively, conducting data triangulation for complicated problems can mean recognizing different causal mechanisms across a range of interventions in different contexts.³⁴

Complex problems refer to interventions where the causal pathway is adaptive or emergent: "...interventions where it is not possible to set out in advance the details of what will be done."³⁵ The intervention may not have pre-identified outcomes, but rather a vague, goal-level description of the desired end-result without a clear pathway of how to get there. It might also include investigation of recursive feedback loops and emergent outcomes, such as unintended changes and resulting system dynamics.

This distinction will help to identify the low-hanging fruit that can be easily measured, and the fruit which requires further and more advanced efforts to reach. Making this distinction may help prepare for the evaluation, such as budget and time allocations, as well as with data collection, tool design and analysis protocols.

Keep in mind that problems are rarely one-dimensional: the simple, complicated and complex programming elements frequently overlap. Measurement efforts can be improved by knowing when and how to simplify, complicate and 'complex-ify' our measurement systems, tools, and approaches.

Patricia Rogers states that "the art of dealing with the complicated and complex real world lies in knowing when to simplify and when, and how, to complicate."³⁶ The key challenge addressed in this paper is to understand social science approaches to measurement problems for simple, complicated and complex results of interventions in fragile and conflict-affected environments.

³² Corlazzoli and White, *Practical Approaches to Theories of Change Part II*, 10-11.

³³ Including monitoring; for further information see, *Interim Guidance Note: Measuring and managing for results in fragile and conflict-affected states and situations*, DFID.

³⁴ Rogers, "Using Programme Theory," 28.

³⁵ Rogers and Hummelbrunner, *On-line e-Learning programme*, 124

³⁶ Rogers, "Using Programme Theory," 28.

Social Science Solutions

We now turn to examining potential social science solutions to measuring the un-measurable. The section is structured around the unit of investigation:

1. Outcomes;
2. Impact and Long-term measurement; and
3. Peace Writ Large

This structure corresponds with what is simple, complicated and complex in programme theory, and associated tools and strategies for overcoming the challenges that each of these classifications might pose.

Each section lists a range of qualitative or quantitative tools and/or methodologies that can be leveraged to more rigorously measure the corresponding unit, provide practical tips for design and implementation, and describe nuances in using the tool.

While this paper is focused on solutions to measurement challenges in situations of fragility and conflict, it is nevertheless important to note what practitioners currently do well in such situations. The following staple data collection tools have been refined throughout the social sciences and these lessons should be taken into consideration: case studies; focus group discussions; key informant and group interviews; observation; participatory approaches; and surveys.

Rigorous literature on best practices and how-to guides abound for these data collection tools and should be utilised to inform the application of focus groups, key informant interviews, and surveys. That said, even when these tools are applied correctly, they might not be sufficient to overcome the challenges of measuring intangible concepts. Combining these widely known tools with the approaches and methodologies outlined in this publication will get us closer to measuring the un-measurable.

The devil, however, is in the details; namely, data availability and the criteria for data interpretation and analysis. Most of the existing literature on the aforementioned tools focuses on best practise in constructing and implementing them. There is a need for equally rigorous and thorough criteria for data interpretation and analysis to be outlined. While it is beyond the scope of this paper, this is an area for further research in to peace & conflict, security & justice programming.

Where possible and appropriate, participatory tools and methodologies should be utilised. This will increase downwards accountability by ensuring that programming is held accountable to programme participants, as well as empower participants themselves to reflect on the conflict and changes taken place.

Finally, no single tool, method or approach is a cure-all to measurement challenges in peace & conflict and security & justice. There is widespread agreement on the value of employing mixed method techniques from a range of disciplines.³⁷ As a recent DFID working paper, *Broadening the Range of Designs and Methods for Impact Evaluations*, stated: "different techniques meet specific purposes, from measurement and description of events and states to understanding of a situation or a process, bringing their own strengths and limitations. Combining methods is a way to overcome limitations and enhance

³⁷ See, for example, Greene, Jennifer C., and Valerie J. Caracelli. "Defining and describing the paradigm issue in mixed-method evaluation." *New directions for evaluation* 1997, no. 74, 1997: 5-17.

strengths.”³⁸ The mixture of tools, methods, and approaches as a means of leveraging the full weight of the social sciences to overcome the challenges outlined above is a recurring theme throughout this paper. Each tool has its own strengths and weaknesses, and it is only by identifying what one needs to know in order to improve programming that a strategy to identify the best tool can be put in place.

PART I: SOCIAL SCIENCE METHODOLOGY FOR OUTCOME MEASUREMENTS

The tools presented below are particularly well-suited for outcome measurement in monitoring and evaluation. They are a combination of qualitative and quantitative tools, some of which are more amenable to participatory processes than others.

The social science tools in this section can be adapted for measuring change at a variety of levels. Many are appropriate at the local, community or district level, such as Community Score Cards (p. 17) while some can provide valuable aggregate data for more accurate macro-level pictures, particularly if standardized across funded programmes, such as proxy indicators (p. 20). Others can be adapted for the micro, macro and Peace Writ Large levels, such as systems analysis (p. 41) and indices (p. 19). For example, an index is a summary or accumulation of scores from a variety of individual indicators that rank-orders specific observations in order to represent a more general concept. As such, we can use the index to look at the micro level by looking at individual indicators or individual countries. We can take it one step further and use the index to look at more macro-levels by comparing across regions and countries. We can even use the index to look at Peace Writ Large by looking at how the index uses the individual indicators as a measure for a larger, more general concept.

All tools will need to be adapted to fit the local context and unit(s) of investigation. No single tool is sufficient to accurately capture the complex realities of change. They must be used in conjunction with one another. The selection of the tools should be informed by their strengths and weaknesses for the intended purpose.

Likert scales

Likert scales are a way to frame questions during surveys, interviews, and focus groups. Likert scales give a more nuanced perspective than simple yes-no questions by providing a range of scaled responses.

Likert scales are useful as both micro- and macro-scales. Generally, data is aggregated within each level to determine the distribution of opinions along the pre-determined scale. In the example below, likert scales were used in a baseline assessment in Indonesia.³⁹ The excerpt shows how four sub-questions were used to measure feelings towards women.

Table 1: Likert Scale Example from SFCG Baseline Survey

In the next set of questions, I would like for you to tell me how much you agree or disagree with each question.			
5.4	Do you believe that women...		
5.4a	...should be able to be the community leader?	Strongly disagree.....1	

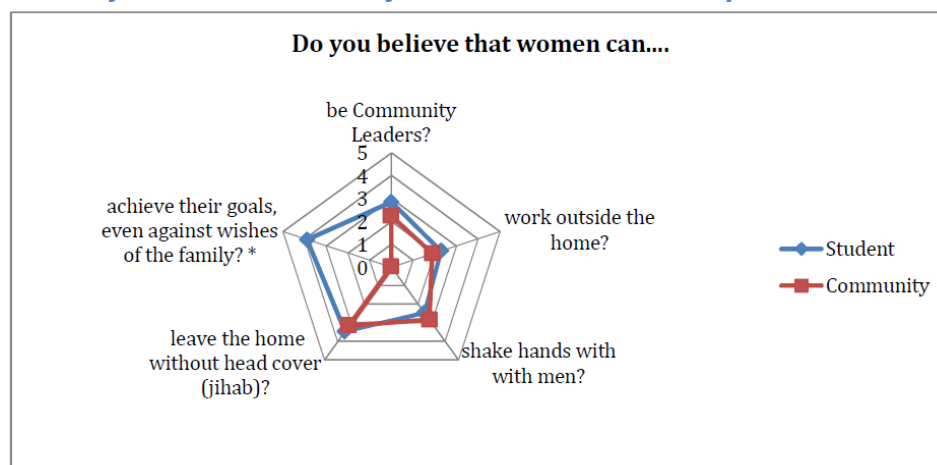
³⁸ Stern, Elliot, Nicoletta Stame, John Mayne, Kim Forss, Rick Davies, and Barbara Befani. "Broadening the Range of Designs and Methods for Impact Evaluations: Report of a Study Commissioned by the Department for International Development." *DFID: Department for International Development*. 2012: 31.

³⁹ Corlazzoli, Vanessa. *Baseline Report: Countering and Preventing Radicalisation In Indonesian Pesantrens*. Washington, DC: Search for Common Ground, 2011.

		Disagree.....2 Neither agree nor disagree/Don't know.....3 Agree4 Strongly agree.....5	
5.4b	...should be able to work outside the home?	Strongly disagree.....1 Disagree.....2 Neither agree nor disagree/Don't know.....3 Agree4 Strongly agree.....5	
5.4c	...should be able to shake hands with men?	Strongly disagree.....1 Disagree.....2 Neither agree nor disagree/Don't know.....3 Agree4 Strongly agree.....5	
5.4c	...should be able to leave the house without their head covered?	Strongly disagree.....1 Disagree.....2 Neither agree nor disagree/Don't know.....3 Agree4 Strongly agree.....5	

The results of the likert scales were displayed using spider charts to offer an easy visualisation of the data, comparing the answers of students with those of community members. The spider chart allows the reader to see the answers to each of the 4 sub questions (plus one extra question that was only asked to students), divided into students and community members. The middle of the 'web' corresponds with 0, with each subsequent strand out representing the next level of the likert scale. The farther away from the centre, the more the respondents agree with the statement.

Table 2: Visualisation of Likert Scale Results from SFCG Baseline Survey



(source: Corlazzoli, *Baseline Report: Countering and Preventing Radicalization in Indonesian Pesantrens*, 14.)

To show a bigger picture, the answers from each community could be compared to those from other communities, in order to 'add up' to show women's status in Indonesian society as a whole. The same questions and likert scales will be used to evaluate the project by comparing the answers from the baseline to those at the end of the project in order to show a change in beliefs, attitudes, and perceptions towards women.

Likert scales provide a nuanced perspective of *subtle* change and the *pace* of change. They can be used in surveys and focus groups to track minor changes in group attitudes towards tolerance of the 'other.' For example, a change from 'strongly disagree' to 'disagree' in regards to the rights of another group to live in the village may be significant in the face of deeply engrained intolerances and the usual short time period given to implementing agencies to achieve such changes in conflict and fragile states. This tool can be particularly useful in measuring cultural or attitudinal change, which is often slow and incremental.

Likert Scales can be limiting in that they cannot explain *why a change happened*. This can be mitigated by adding an open-ended question asking the respondent to explain the rationale for their choice. The Likert scale can be easily combined with other tools, such as focus groups, for standardising responses. Likert data can be easily quantified and analysed, such as using statistical analysis.

It is important to remember that Likert scales are generally seen as ordinal and not interval scales.⁴⁰ This means that the difference between "5-excellent" and "4-good" may not be the same as the difference between "4-good" and "3-average," despite both groups being scored as one point away from each other. Some scholars hold that one may be able to draw conclusions based on the percentages of responses for each scale (i.e.: when asked about the quality of the training, 56% said it was excellent), but may not be able to average the entire scale (i.e.: the training was given a quality rating of 3.7) due to its ordinal nature.

Nevertheless, it may be possible to draw more rigorous statistical conclusions using 7-point scales, as opposed to the common 5-point scale. Doing so may mitigate, to an extent, Type I and II statistical errors commonly associated with departures from the intervals.⁴¹ Such a method can be used to, for example, examine aggregate inter-group attitudes towards the 'other' as a means of tracking outcome and impacts on inter-group knowledge, attitudes, behaviours, and perceptions. In this way, Likert scales may be used for complicated problems.

Building evaluative terms into the scaled responses can increase the utility of the data.⁴² For example, respondents are often asked their extent of agreement on, for example, quality. The utility of such data is limited because it is about agreement not the respondents' judgement of quality.

Community Score Cards

Community Score Cards (CSC) are a quantitative, participatory tool most often used to solicit community members' "*perceptions on quality, efficiency and transparency*" of community service providers and their performance at the *local level*.⁴³ Service providers may include implementers, donors, companies,

⁴⁰ Jamieson, Susan. "Likert scales: how to (ab) use them." *Medical education* 38, no. 12, 2004: 1217-1218.

⁴¹ Jaccard, James, and Choi K. Wan. *LISREL approaches to interaction effects in multiple regression*. Vol. 114. Sage Publications, Incorporated, 1996.

⁴² Davidson, Jane. "Don't expect quantitative evidence to answer a qualitative evaluation question." *Genuine Evaluation*. <http://genuineevaluation.com/dont-expect-quantitative-evidence-to-answer-a-qualitative-evaluation-question/> (accessed March 14, 2013).

⁴³ Singh, Janmejaj, and Parmesh Shah. "Community Score Card Process: a short note on the general methodology for implementation." *Social Development Department, World Bank*, 2003: 1; italics added.

or local institutions, such as police departments, government and district officials, and judiciary procedures. CSCs provide a mechanism for actors associated with an intervention to receive feedback on their behaviour, attitude, or conduct. CSCs may also be used as a programming tool.⁴⁴

This could be particularly useful for assessing sensitivity to the local dynamics of conflict and/or fragility. For example, CSCs could reveal that services are only being provided in one part of a community, thereby exacerbating local tensions. It may highlight criterion for determining the award of future funding for individual projects or programmes in that region. As such, it may act as a tool for remote monitoring (p. 29). For instance, a donor could use the results of CSCs to ensure that programmes are being implemented as intended by directly consulting the beneficiaries. If adequately created and administered, this tool can promote and empower local perspectives and capacities, while still meeting the need for mass aggregate data for donor or agency needs.

Furthermore, CSCs assist in identifying incremental changes towards objectives within targeted groups or sub-groups. They can be particularly useful in listening to perspectives within the community and identifying 'the path of least resistance' towards the desired changes (defined or emerging). For example, a programme focusing on improving accountability and transparency could use CSCs to identify key government officials to work with. At the same time, the identified incremental changes can lead to the identification of plausible theories of change to achieve the objectives. This can serve as the basis for future activities, interventions and/or funding opportunities.

CSC data can also be re-used, when relevant, as baseline data or in the programme design stage. For baselines, CSC data can feed into upcoming initiatives by providing key perception measures on a wide range of issues, such as effectiveness of government. By using this data as a baseline, programme managers are able to set more realistic and achievable targets and benchmarks. Similarly, such data can inform the choice of indicators for programmatic outcomes or targets, such as identifying what 'success' on a particular topic might look like from a local perspective. Taken a step further, the data can identify potential outcomes that programmes might seek to work towards in that community or specific context.

CSCs may also be used to monitor whether critical assumptions (either at each level of the design hierarchy or its overarching theory of change) have held true in the local context at the micro scale. For example, in a hypothetical project, a key assumption might be that an increase in the number of women holding seats in the local government would lead to an improvement in women's rights in the community. A CSC could ask community members about their perceptions of women's rights. This might reveal that women have been included in leadership positions without being extended any real power. To do so, researchers will need to take into account issues such as culture or conflict and fragility dynamics. Using CSCs at the national or macro-scale is generally not recommended due to the degree of facilitation, mobilisation, follow-up, and analysis required.

⁴⁴ CSCs can be used to collect key information from the community to obtain accountability and responsiveness from service providers. It is intended to produce improvements in public services or overall efficiency of service delivery. Keep in mind that the CSC itself will not improve delivery: it provides key information to determine the strategy and tools to improve service delivery. It is essential to take into consideration local incentives and power dynamics in governance and other areas that might affect service delivery. These could include the extent to which local official buy-in is required, or what the local customs are for achieving such buy-in.⁴⁴

Example 1: Community Score Card use in Malawi

Community Score Cards (CSCs) have been increasingly used in DFID-funded programs. In **Malawi's** Community Based Monitoring Programme, CSCs were employed to determine why a governance programme focusing on improving the delivery of public services had been more effective at the local level than it had been at the national level.

Data collection was facilitated in local languages and the CSCs included pictures to ensure marginalized and illiterate populations were reached. Program staff felt that CSCs helped improve communication between community members and service providers, including local government officials, empowered local communities' own capacity for self-help, and reduced corruption in the distribution of agricultural coupons.

The evaluators noted that challenges of scaling-up could be mitigated by including stakeholders from the state and national government levels. Overall there was a need for increased capacity building at the local level, which could be mitigated by more intensive training of local facilitators on the tool.

(Source: Wild and Harris, *More than just 'demand'*)

Regardless of the purpose or intended use of the CSC and its findings, it is essential to ensure participant expectations are realistic. Experience⁴⁵ demonstrates that if the purpose of this tool is not carefully explained, participants might take the exercise as an indication of impending assistance on the issue(s) at hand. It is therefore essential that the purpose of the research is clearly outlined to participants, along with how the results will be used. Transparency to participants in the purpose and use of research is a fundamental principle of conflict sensitive research in fragile and conflict-affected environments.

Indices

An index is a summary or accumulation of scores from a variety of individual indicators that ranks specific observations in order to represent a more general concept. An index is often the end result of a survey. For example, UNDP's Human Development Index (HDI) ranks countries into levels of human development by collecting and analysing indicators such as life expectancy, education, and income. Other well-known indices include the World Bank's World Development Indicators and Afrobarometer. For a comprehensive list of indices relevant in FCAS, see Annex A.

Indices can provide a valuable resource-saving way of collecting data on key, standardised indicators across the spectrum of international assistance. The collection and centralisation of such data has numerous benefits, as a recent DFID paper outlined: the "production of data is continuous, allowing for better analysis of trends; locally-situated and managed, allowing for capacity transfer between external and local partners; and multi-agency, supporting the involvement of multiple government, NGO and international organisation partners."⁴⁶

For example, the UNDP's Human Development Index⁴⁷ shows how to use an index to measure different levels of change. The HDI collects data that supports various indicators of human development. At the micro level, the UNDP provides indicators on life expectancy at birth, mean years of schooling, expected years of schooling, and Gross National Income (GNI). These indicators are chosen to be broadly

⁴⁵ See, for example, Addai, Emmanuel, Emma Kpenu, and Martin Dery. "The Community Scorecard Approach for Performance Assessment: ProNet North's Experience." *WaterAid Ghana*, no. 4, 2004.

⁴⁶ Taylor, Sebastian. *Practical Approaches to Strengthening Data and Evaluation in Fragile and Conflict Settings*. theDLgroup, DFID Somalia, 2012: 31.

⁴⁷ *Indices & Data: Human Development Index*. Human Development Reports, United Nations Development Programme (UNDP). <http://hdr.undp.org/en/statistics/hdi/> (accessed March 28, 2013).

informative about the quality of life within individual countries. UNDP also allows the user to rank countries according to a selected indicator in order to support cross-country or cross-regional analysis.

Another widely used index is the Failed State Index, which is published annually by Foreign Policy and the Fund for Peace. This index offers a user-friendly interactive world map using colours to highlight aggregate measures of state failure. The index ranks countries according to 12 indicators and various sub-indicators which relate to social, economic, and political drivers of conflict and fragility. These drivers include demographic pressures, levels of human flight, inequality, corruption, human rights abuses, access to improved services, presence of ethnic violence, group grievances, levels of authoritarianism and foreign assistance. Users can rely on these indicators for baseline data, triangulation, and other analytical purposes in their programmes.

The increasing publication of localised data indices, such as the Liberia Armed Observatory or the Jamaica Crime Observatory, represents a promising trend for data analysts.⁴⁸ The emergence of additional local data observatories will strengthen the overall global network of data, as well as 'go deeper' into more localised indicators relevant to the primary purpose of the index. As DFID notes, the observatory model holds promise for systems analysis (p. 41) and knowledge generation in general, but particularly for inputs to "planning and intervention design as well as backstopping for the evaluation of impact."⁴⁹

Locally-led data collection and subsequent centralisation can be utilised in the re-construction of baseline data. It also offers a way to access locations where it may be difficult to gain regular access. Practitioners could consider increasing local data collection capacity in terms of peace & conflict and justice & security by encouraging national observatories to weave one or two additional relevant questions into their data collection frameworks.

It is important to examine the sources and methodologies used by each index to make sure they are appropriate for the desired research. Some indices only use secondary sources, such as surveying media content (p. 24), others use primary data collection methods; some update their content regularly, others annually. It is important to consider that indices may have different data sources for the same indicator, in which case DFID advises using data from the UN Statistics Division.⁵⁰

Proxy Indicators

Proxy indicators are used to approximate the reasonable likelihood that a change occurred when direct measurement is not possible.⁵¹ They are distinct from traditional indicators in that traditional indicators are direct measurements of change, while proxy indicators are representations of measurable changes that suggest that a broader, unmeasurable change (i.e. outcome) has occurred.

According to UNDP's definition, proxy indicators do not indicate that change has occurred but rather suggest that there are *conditions conducive* to the desired change.⁵² This can be particularly useful in fragile and conflict-affected environments, where regular data collection mechanisms are insufficient or non-existent, or where certain lines of inquiry may be inappropriate (for example, on gender-based

⁴⁸ Taylor, *Practical Approaches*, 32-33.

⁴⁹ Taylor, *Practical Approaches*, 32.

⁵⁰ *Interim Guidance Note: Measuring and managing for results*, DFID, 13.

⁵¹ Majchrzak, Ann. *Methods for policy research*. Vol. 3. Sage Publications, Incorporated, 1984: 57.

⁵² United Nations Development Programme. "Proxy indicator." *Monitoring and Evaluation: Project framework, baselines, indicators and targets*. <http://hrba.undp.sk/index.php/monitoring-and-evaluation/project-framework-baselines-indicators-and-targets/proxy-indicator> (accessed March 21, 2013).

violence or in relation to a recent traumatic experience). Proxy indicators also offer a way to measure more abstract concepts, such as culture, trust, justice, and reconciliation.

An example of using proxy indicators is to assess the change in the institutional structures that underpin the distribution of power between the state and societal actors in one single measurement or question. Since this is not directly measurable, proxy indicators could be used, such as % of seats in parliament held by women or % of seats in parliament held by minority ethnic groups.

Proxy indicators are based on assumptions that a measurable item is indicative of the desired result. It is essential that these assumptions are a) few in number, and b) logically sound and realistic.⁵³ This can be done by identifying the result the proxy indicator measures and the assumptions on how the desired result will lead to the observable proxy. The indicator will be weaker when more assumptions are required to arrive at the proxy. The following formula may be used:

$\text{result} + [\text{assumption}] = \text{proxy indicator.}$

For instance, our desired result might be to increase access to justice, and one of the proxy indicators used is # of new courts opened. This assumes that physical access to courts was the key impediment for increased justice.⁵⁴ Keep in mind that context and culture matter, and what might be a reasonable assumption in one context may not be true in another. For this reason, it is recommended that one use multiple proxy indicators to measure a single result.

Many scholars and practitioners advocate using groups of indicators, sometimes referred to as *baskets* or *bundles of indicators*. For instance, a recent DFID paper stated that it is “critical that the indicators collaborate, that they are a bundle that collectively paints a rich picture of how change happens and what change looks like.”⁵⁵ Using bundles of related proxy indicators will help measure the un-measurable by looking at the nuances of change.

Table 3: Example of DFID Governance and Conflict Bundles of Indicators

Peace Process Support <ul style="list-style-type: none"> • Cross check with Elections, Political Party Support 	Purpose/ Outcome	% negotiating partners publicly advocating peace through promotion of next plausible steps in peace process
		% citizens content with progress of peace process
		% of terms of peace agreement implemented successfully according to independent assessment
		# ceasefire violations (if ceasefire in place)
	Output	% targeted citizens aware of peace process
		% peace process mediators trained to international standard in conflict resolution
		% negotiations facilitated by negotiation team
		% drivers of conflict included in negotiations
		Rating of extent to which negotiation plan includes multiple approaches to influencing and a range of workable solutions

(source: Barnett et al, *Governance and Conflict Indicators Report*, 66.)

The triangulation of fact-based proxy indicators is essential for perception-based indicators. Perception indicators are subject to interpretation and not based on empirical evidence (i.e. fact-based indicators). As previous DFID guidance documents have stated: “Perception indicators are important... but they

⁵³ *Performance Monitoring & Evaluation TIPS: Selecting Performance Indicators*, 2nd ed. Washington, DC: USAID: United States Agency for International Development, no. 6, 2010.

⁵⁴ *Performance Monitoring & Evaluation TIPS: Selecting Performance Indicators*, USAID, 5.

⁵⁵ Barnett, Chris, Julian Barr, Angela Christie, Belinda Duff and Shaun Hext, *Governance and Conflict Indicators Report*, ITAD and DFID: Department for International Development, 2011: 23.

often need to be balanced with fact-based proxy indicators to gain substance.”⁵⁶ For example, if a perception-based indicator is % change from baseline of respondents who state local authorities are responsive to lodged complaints, then an associated proxy indicator might be change in ratio between # of documented responses to complaints. Neither is immediately indicative of greater accountability of governance structures, but in combination provides evidence that such accountability may be increasing.

Information and Communications Technologies

Information and Communications Technologies (ICTs) hold utility in fragile and conflict-affected states for their ability to produce large data sets, which can be combined with other tools for greater analytical capacities and insights. New technologies have the potential to reach the un-reachable, if certain conditions in the country exist.

In particular, ICTs hold promise in their ability to reach across geographic scales regardless of the state of the context (i.e., an increase in violence in one area making it impossible to access others) and help in reaching the un-reachable, voiceless and/or marginalised populations. Their strengths also include:

- Real-time automated data aggregation and analysis through software.
- “Improved adherence to complex or context-dependent questionnaires, as the device determines which questions should be answered or skipped.”⁵⁷ For example, validation checks can be prepared in the software, as can logical question flow, automated data cleaning and real-time data quality checks.⁵⁸
- Reduced data collection costs and time resources. This is particularly useful for large-scale surveys and other research methods.
- Collection of additional data types, including GPS locations.
- Use of multiple languages and communication methods, such as pictures, that can reach illiterate and disabled populations.
- Ease of pairing with other tools and strategies described in this document, such as remote monitoring.

Simple phones are defined as basic phones capable of phone-calls and SMS, but not internet connectivity or GPS. Data can be collected through SMS or a phone-call to an operator to provide answers on the lines of inquiry. These are primarily used for relatively simple surveys, questionnaires, and short qualitative responses.

SMS can be employed as a data collection tool by formal or informal enumerators or researchers, or by the respondents themselves. As a tool, it is capable of accommodating mandatory ratios for respondents, such as gender, age, or otherwise. These requirements can often be built into the software used to manage the data collection, such as FrontlineSMS.⁵⁹ The accuracy of the SMS data may vary depending on the protocols and software used.

In contrast to the automation of SMS, phone calls allow for real-person contact and greater complexity in the lines of inquiry or the format used (key informant interview, survey, etc.). This may cost more

⁵⁶ Barnett, Barr, Christie, Duff and Hext, *Governance and Conflict Indicators Report*, 23.

⁵⁷ Patnaik, Somani, Emma Brunskill, and William Thies. "Evaluating the accuracy of data collection on mobile phones: A study of forms, SMS, and voice." In *Information and Communication Technologies and Development (ICTD), 2009 International Conference on*, pp. 74-84. IEEE, 2009.

⁵⁸ *Mobile-Based Technology for Monitoring and Evaluation*. Clear Initiative Regional Centers for Learning on Evaluation and Results, 2013: 8-9.

⁵⁹ "FrontlineSMS." FrontlineSMS. <http://www.frontlinesms.com/> (accessed March 29, 2013).

than SMS, but varies from context to context. SMS data can be supplemented by follow-up phone calls to allow for more depth.

Example 2: ICTs in SFCG Programming

Most of Search for Common Ground country programmes use ICTs to monitor and evaluate their programming. SFCG's Promoting Inclusive and Participatory Elections (PIPE) project in **Sierra Leone** used FrontlineSMS, a free SMS database software, to create a database of telephone numbers and send SMS "blasts," a function that allows SFCG to send a single text to multiple recipients with one click on the computer. In **Nigeria**, FrontlineSMS and a call-in hotline have been used in the Niger Delta to record listener feedback about peace and reconciliation radio dramas. Likewise, programming in the DRC has used SMS and a hotline to solicit listener feedback, as well as market other peacebuilding radio programs through SMS blasts.

Use of ICTs allows staff to devote a standard line for texting and calling, maintain a database that permanently archives messages, and offers beneficiaries a cheap, anonymous way to provide feedback. A recent SFCG research project found that higher participation rates in **Liberia** in SMS-based data collection, compared to other countries analysed, was partially due to better technology set-ups. Liberian radio stations had set-up partnerships with private phone companies to ensure constant electricity and provide a choice of networks to allow people to participate at a reduced, or even free, rate. This same finding held true for Ushahidi's experience monitoring Liberia's 2011 elections, where they negotiated for free short codes from three major cell phone network providers allowing Liberians to report irregularities without incurring personal costs.

Many of SFCG's country offices have noted technical challenges and lack of technical capacity and know-how. Electricity cut offs, computer breakdowns, station generator blowouts, and complicated technology systems all affect the ability to collect data.

(Sources: Oliphant, *Desk Review for Participatory Early Warning*, 6-8; Chungong et al, *Community Radio, Gender & ICTS in West Africa*, 1-17; *Empowering the Edge: Information Sharing in Post-Civil War Liberia*, The GIS Professional.)

Smart phones and tablets, on the other hand, have broader applications due to their impressive and increasing computational abilities, including: the use of complex forms, questionnaires and surveys; key informant interviews or focus group discussions; the ability to transmit data in computer-friendly formats (PDF, XLS, RSS, etc.); record locations (GPS); and the ability to connect to Wi-Fi internet in the absence of cell signals. They can be leveraged for the use of video in data collection, which is explored in a later section on Diaries and Video Logs (p. 26). It is beyond the scope of this guidance to provide in-depth details on how to use each of these applications.

The increasing use of GPS-based data through smartphones has helped many actors better understand information and track changes over time and space. For example, with geographic information systems (GIS) it is possible to analyse whether programmatic outputs, such as roads, have had a statistical effect on incidents of violence due to proximity and/or access to resources. This information can be displayed visually on a map, such as with ArcGIS⁶⁰ or Ushahidi⁶¹ software, to deepen analysis and provide alternative perspectives on the data through its visualisation. In this way, new patterns, dynamics and issues may arise that would not have been otherwise observed.

There is, however, a word of caution when employing ICTs for data collection: the tool must be appropriate for the task at hand. Technology is not a cure-all to the challenges of data collection in fragile and conflict-affected environments. The strength of the data and overall research process can be significantly increased by deploying ICTs in conjunction with other data collection tools and methods.

⁶⁰ "ArcGIS - Mapping and Spatial Analysis for Understanding Our World." Esri - GIS Mapping Software, Solutions, Services, Map Apps, and Data. <http://www.esri.com/software/arcgis> (accessed March 21, 2013).

⁶¹ "Ushahidi." Ushahidi. <http://www.ushahidi.com/> (accessed March 21, 2013).

Incident Log

Incident logs refer to the systematic recording and, to an extent, investigation, of certain incident types. It aims to measure the number and gravity of incidents that took place in a given area or time frame. It is a simple tool that can be used in data triangulation and for the complicated measurement of the causal relationship between an incident and cultural attitudes.

At its most basic level, incident logs are a standard tool for programmes operating in fragile and conflict-affected environments and offer a means of tracking, both for monitoring and operational security purposes, the number of violent incidents in the implementing area. In some instances, the very use of incident logs may have outcome-related effects in the environment. For example, tracking incidents of gender-based violence may result in an increase in the number of reported cases because the very presence of the incident log and accompanying activities may have increased trust between citizens and local institutions. The typology of incidents may include disaggregation based on which actors were involved and the gravity of the incident.

Incident logs can also be used to track the relation and effect of incidents on popular attitudes, under the assumption that there is a high correlation between relevant attitudes and the incident type in question.⁶² In the above example, the change in popular attitude was the increase in trust of local institutions for the reporting and redress of gender-based violence. In this way, incident logs can be used to track cultural changes provided there is a high correlation between the incident type and popular attitudes.

Incident logs can also track the ways in which funded interventions are being affected by the environment. For example, DFID Nepal used incident logs in grantee quarterly reports to track “the extent to which their [implementing partners’] activities have been affected by the security situation (e.g. implementation rates or ability of staff to travel within districts).”⁶³ This data could, for example, be input into systems-based analyses (see p. 41) to understand the extent to which a system might be in a reinforcing or counteracting causal loop. An example of two simple causal loops is shown in systems-analyses (see pg. #). Incident logs can be used in constructing these kinds of loops by giving additional data points to be included in the causal loops, making them more comprehensive. While a single incident may not be reason for concern, a series of incidents can indicate an increase in violence.

Incident log data can help inform future funding priorities by establishing patterns on particular incident types and attitudes over-time. However, attaching funding to incident logs may lead to over reporting or over exaggeration of incidents. Longitudinal data of incident logs can be used in later studies, such as impact evaluations (p. 35) and longitudinal studies (p. 43).

Media Content Analysis and Discourse Analysis

Media content analysis is a mixed-methods approach that seeks to describe “with optimum objectivity, precision, and generality, what is said on a given subject in a given place at a given time” in or by the media.⁶⁴ It includes analysis of “media coverage, placement of stories, tone, and visual images, prominence of quotes/personalization, and reach of a media outlet” in order to track how media outlets

⁶² *Violence against Women and Girls: CHASE Guidance Note Series: A DFID Practice Paper*. DFID: Department for International Development, no.3, 2012: 22.

⁶³ *Working Effectively in Conflict-affected and Fragile Situations Briefing Paper I*, DFID, 6.

⁶⁴ Macnamara, Jim. "Media content analysis: Its uses, benefits and best practice methodology." *Asia Pacific Relations Journal*, vol. 6, no. 1, 2005: 1-34.

cover various subjects, issues, or themes.⁶⁵ Media may refer to print, radio, television, websites, or any other form of social media. Practitioners and scholars suggest mixed-methods approaches to media content analysis offer the rigorous methodology of quantitative methods, while still capturing the media content's "likely meanings to and impact on audiences" through qualitative methods.⁶⁶

Media content analysis is applicable beyond media-based programmes. For example, impact level indicators for a programme that seeks to train government officials in security sector reform may include an analysis of media discourse to see if any wider changes have occurred as a result of that government-level training. This might include government speeches or interviews given to the media by the trained officials. The findings of the analysis can be compared to previous speeches to help determine the impact of the programme or activity. Similarly, a consistent shift in the way in which a certain subject is framed, approached or discussed may indicate wider cultural changes.

Qualitative approaches to media content analysis examine the relationship between the media content and its likely audience in order "to determine the likely meaning of texts [or other media content] to audiences."⁶⁷ It pays particular attention to audience characteristics and media and contextual factors that may influence how content is interpreted. In other words, it seeks to understand *how audiences are likely to react to media content* in terms of knowledge, attitude, behaviour and perceptions: the un-measurable changes this guidance is centred upon.

A sub-methodology to qualitative approaches to MCA is discourse analysis. It examines the key assumptions underpinning the discourses that influence social constructions and interactions between people by identifying one or all of the following:

- What the discourse implies;
- What discourse does or what is done with it, and;
- What the discourse says or what wants to be said within discourse.

In this way, discourse analysis helps measure the un-measurable by analysing the relationship between discourse and societal effects. Paired with other methodologies, such as systems analysis (p. 41) or social network analysis (p. 30), this method can illuminate key relationships and variables in discourse that influence effects.

Quantitative approaches to media content analysis collect data on topics, volume of mentions, audience reach and frequency, and the messages being conveyed by key words in context.⁶⁸ A quantitative approach is particularly conducive to understanding *changes in quantity and reach* of particular media outlets or contents (i.e., a particular phrase, approach to a subject, etc.). If paired with qualitative analysis, this method may be used to identify which media outlets are contributing positively or negatively to societal knowledge, attitudes, behaviours, and perceptions. This in turn can be used to prioritize intervention target groups, regions, topics, etc.

⁶⁵ Sartorius, Rolf, and Chris Carver. *Monitoring, Evaluation, and Learning for Fragile States and Peacebuilding Programs: Practical Tools for Improving Program Performance and Results*. Social Impact: 94.

⁶⁶ See Macnamara, Media content analysis, 1-34 for an in-depth analysis and literature review of the quantitative vs qualitative MCA debate.

⁶⁷ Macnamara, Media content analysis, 5.

⁶⁸ Macnamara, Media content analysis, 4.

Example 3: Media Content Analysis in Yemen

Media Content Analysis (MCA) has been successfully employed by DFID funded projects with BBC World Service Trust (now, BBC Media Action) to assess the impact of a journalist training programme on the standard of reporting for Yemen's 2006 local and presidential elections. Media Content Analysis was used to triangulate data from surveys, interviews, and a review of project documents, and was particularly aimed at overcoming self-reporting bias of these other tools. MCA proved useful for measuring change in media output, and detected changes in content, presentation, and production aspects. Researchers noted that MCA alone could not show attribution, or even the effect of the training on trainees, and needed to be used in conjunction with other methods to reveal the whole picture.

(Source: Whitehead and Saville. *Using content analysis to measure the influence of media development interventions: Elections training for journalists in Yemen.*)

Keep in mind that predetermined criteria for analysis and interpretation of media content is required to ensure that data collectors are all using the same definitions to categorize content. The need for common criteria transcends across all approaches to media content analysis.

Media content analysis can be used as an analytical approach to media outputs and, more broadly, as a means of assessing the *presence of enabling or conducive environmental conditions for perception, knowledge, attitude, and behavioural change*. It allows for the analysis, human or computerised, of a wide range of data over time to identify popular discourses and their likely meanings and implications for societal change. It may occur on a frequent, recurring basis as an alternative to expensive, large-scale audience research initiatives, which are often restricted to year-based timeframes.⁶⁹

Participant Diaries and Video Logs

Participant diaries and video logs are participatory qualitative research methods.

Diaries, written or recorded (video or voice), are primarily oriented towards learning about knowledge, attitude, behaviour, and perspective changes as they occur from the perspective of the individual participant. Both tools empower the participant to reflect upon, identify and share significant experiences and perspectives relating to the changes they have experienced or the changes they witness in their immediate surroundings. Participants use their own words and ascribe meaning to the changes they witness.⁷⁰

Video is particularly well-suited for in-depth non-verbal communication, analysis and observation. It is a useful tool that allows illiterate populations to express themselves quickly and accurately. By working with technology, the beneficiary may also learn new skills.

Written diaries are particularly well-suited to elicit sensitive information that participants may not feel comfortable sharing or disclosing openly in a public setting or with others.⁷¹ Depending on cultural and individual preferences, writing down one's thoughts may be a strange concept. If this is the case, video may be a better option.

Information obtained through video or diary can be used by the monitoring or evaluation team to interpret the meaning behind the attitude or behaviour in the video or in the diary. The information

⁶⁹ Macnamara, Media content analysis, 8.

⁷⁰ Day and Thatcher, Reflections on using diaries, 249-259.

⁷¹ Day, Melissa, and Joanne Thatcher. "I'm really embarrassed that you're going to read this...": Reflections on using diaries in qualitative research." *Qualitative Research in Psychology* 6, no. 4, 2009: 249-259.

may be analysed to understand why certain change(s) occurred.⁷² The participant might even analyse themselves the content of the diaries to identify changes over time. By watching older videos of themselves, participants have an opportunity to agree or disagree with statements they made previous to the intervention, and explain why and how their views evolved further. This would strengthen the reflection and empowering aspect of utilizing diaries.

Using video or diaries over a long period of time will strengthen the rigour of the data collected. For instance, participants can be asked to produce videos before, during, and after a programme. The sequence of the videos can show if there was a change in confidence, views, and attitudes towards a conflict. This method could also be used to learn about individual level changes, and how these have interacted positively or negatively with shift in the conflict or in the culture.

Diaries can also be used as monitoring strategies to identify when a concept or a shift in attitude actually takes place within an intervention. The video may be able to show which a particular training session resonated with a participant. It may also bring to the forefront how previous micro-changes may have added up over time. Stronger conclusions on the effect of a programme may be drawn if multiple participants shift perspectives at a similar point in the intervention or during the same training session.

By combining the direct review of a participant's diary with the review of a pre-set list of behaviours, attitudes, and words, the M&E team will be better able to interpret the diary content and potentially even quantify the qualitative data. For instance, the M&E team can analyse how trained youth used cooperative language in the videos before the training vs. after the training. The video diary can also show improved dialogue skills. For instance, participants can be given the task of conducting a media report before and after a training on best practices in journalism. Ultimately, diaries are a tool that can be used to measure a participant's level of confidence, assertiveness, and empowerment, particular with marginalized populations and girls and women.

The video diaries have some inherent weaknesses. It may not be possible to use this tool to assess changes at the group-level or at the national-level, given that it would require a large amount of resources to collect and analyse all the footage. That said an intervention could ask the greater population to upload video diaries to Dropbox⁷³ and then use a viewer's rating system to analyse the videos. Data-overload and time to analyse the content of the videos or written diaries will increase, if clear boundaries of reflection are not well defined. For instance, the M&E team may consider restricting the length of the video or written diary. However, these boundaries and limits may limit the possibility of discovering unintended, unanticipated results, as well as the scope of the meaning the individual ascribes to the changes. It is essential that the use of diaries include a clear plan from the start on what data needs to be gathered and how it will be used.

⁷² *Understanding Political Violence among Youth: Evidence from Kenya on the links between youth economic independence, social integration, and stability.* Mercy Corps, 2011.

⁷³ No matter where you choose to store your data, it is important to make sure that it is a safe and secure method in order to protect participants and researchers.

Example 4: Video Diaries in Uganda

War Child Holland, an international NGO that seeks to protect children from violence by offering psychosocial support and education, has used participatory video and video diaries to monitor and evaluate their programmes in Northern Uganda.

Project staff feels that video diaries have enhanced accountability and transparency by allowing beneficiaries to communicate directly with their audience (WHO WAS THE AUDIENCE?), yielding insight into micro-level changes. When the videos are analysed all together, these micro-level changes-- in perceptions, attitudes, beliefs, and knowledge-- can add up to more macro-level change.

One of the major benefits of video diaries *"is that there is a whole lot of data that [the beneficiary] can give me, that they can share"*, but that the evaluator may not necessary know to ask or that it exists. War Child Holland has developed an instructional video about their use of video diaries in monitoring and evaluation, which can be accessed here:
http://www.youtube.com/watch?feature=player_embedded&v=srGNFkztrPU.

(sources: War Child. War Child; Raftree, *Using participatory video for monitoring and evaluation*)

As participatory methods, diaries and video logs require significant buy-in and trust from the participants, and can be paired well with assessments of programme outcomes. They are particularly well suited for monitoring strategies as well as evaluation approaches, such as empowerment, outcome mapping (p. 32) and most significant change (p. 39). Keep in mind that this method requires significant time both by the participant and the researcher throughout the process. Safety of the participant must be assured throughout the life of the project, as well as after the data has been collected. In today's ever evolving world of social media, including Youtube, it is even more imperative that all ethical research guidelines be discussed, implemented and ensured to guarantee the safety of the participants, particularly when working on conflict and fragile states.

Rapid Assessment Procedures

Rapid assessment procedures (RAPs) are a mixed-methods approach to action-research that is particularly useful under resource constraints, such as time and money.

It is primarily a qualitative approach that relies on focus groups, key informant interviews and short qualitative survey tools in order to gain a snapshot of a complicated situation, either in general or on pre-determined lines of inquiry. For example, RAPs have been used to collect information on conflict dynamic perspectives from IDPs or refugees. They could also be used to assess service delivery performance and relevance in the face of changing environmental conditions. They may be used to update a conflict assessment or to assess a situation after a violent incident.

RAPs are simple tools that are relevant when there is an evolving environment, including major shifts in dynamics, and developmental programming, as well as for the assessment of quality of programming and/or service delivery. There are many useful guidance notes and resources for RAPs that can be adapted from the humanitarian and food security sectors.⁷⁴

⁷⁴ Freudenberger, Karen Schoonmaker. "Rapid rural appraisal (RRA) and participatory rural appraisal (PRA): a manual for CRS field workers and partners." *Baltimore: Catholic Relief Services*, 1999; Gittelsohn, Joel. *Rapid Assessment Procedures (RAP): ethnographic methods to investigate women's health*. International nutrition foundation (INF), 1998.

Remote Monitoring

Remote monitoring is used in two primary ways. First, remote monitor is used to collect data in hard-to reach locations. It is often used in places where it is difficult or not secure to send enumerators or field-staff. For instance, remote monitoring and triangulation was often used in Afghanistan at the height of the conflict. Local focal points were identified throughout the country, and they collected information and reported back to a centralized location.

The second way that remote monitoring works is to provide aggregate data on standardised indicator sets across programmes, regions, countries, and portfolios. Remote monitoring might seek to track one of the following:

- The evolution of the context, conflict and/or context monitoring, with a specific emphasis on key 'turning points';
- Causation between events that took place and effects they had on the evolution of the context or conflict; and
- The preliminary results of funded programmes in the chosen geographic area, which may be inaccessible to evaluators or researchers.

Some early-warning systems aim to utilize aspects of remote monitoring systems by having key focal points within one region feeding information to a centralized location regarding key indicators related to conflict. Search for Common Ground and local partner organizations have set up an SMS conflict assessment system in Uganda and Nigeria and are testing different models.

Remote monitoring allows for the aggregation of local trends and dynamics across funded interventions. The challenge is to collect accurate data on a timely basis in complex environments. Incentivizing focal points to report on key information over long periods of time is another challenge. Finding ways to safely transport the information and ensure the safety of the focal points is key another obstacle to plan around in remote monitoring. If these challenges are overcome, it is possible to develop a quicker and more responsive understanding of the evolving context.

Overall, remote monitoring pairs well with indices (p. 19), incident logs (p. 24), and proxy indicators (p. 20). The reliance on well-defined standardised indicators is a central piece of remote monitoring.

Example 5: Remote Monitoring in Somalia

Oxfam uses a remote monitoring and remote partnership approach in its programming in Somalia. Since 2007, they have trained 28 local Somali organizations on a variety of topics, including how to conduct RAPs, context analyses, and conflict sensitivity. They focus on M&E at four levels: community, partner, external monitoring agency (chosen from a Somali agency), and Oxfam International.

Oxfam uses a variety of tools to overcome distance and remote monitoring challenges, including: GPS (for example, to identify project sites); mobile phones (calls with local beneficiaries and other stakeholders); photographs (to document project activities, purchases, and services); and SMS-based surveys (primarily aimed at measuring community perceptions about aid and its delivery in Somalia).

Some argue that while international actors frequently question the capacities of local actors, experience has shown that within the Somali context, local actors have proven more capable than their international counterparts at collecting data in their own communities.

(sources: Kamalingin and Noor, *Oxfam's remote partnerships, monitoring and evaluation mechanisms in Somalia*; Tsitrinbaum, *Humanitarian Partnerships Under Fire: A Case Study of Somalia*.)

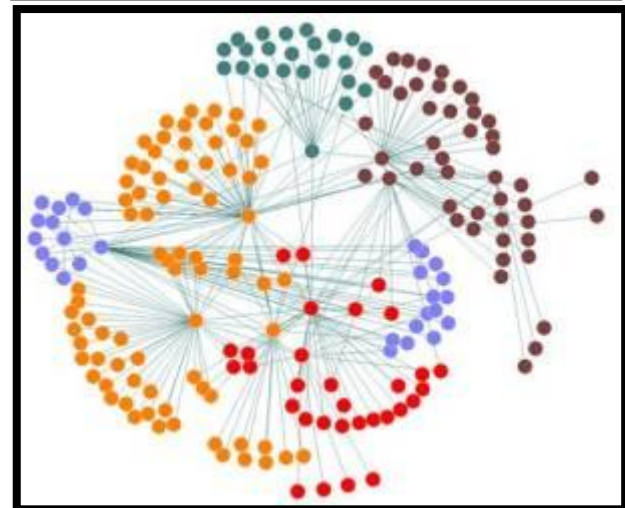
Social Network Analysis

Social network analysis (SNA) is a methodology used to examine human behaviour and social change by analysing patterns of relations and relationships between individuals, groups, and/or organizations. It enables to analysis of the social structures within a particular society or community. Social network analysis views social relationships in terms of a 'network theory' made up of *nodes* (representing individual actors or groups within a network with a point) with *ties* (representing the strength of the relationship or association with a line).⁷⁵

Social network theory holds that social change is generated by the relationships between actors and their interdependence.⁷⁶ As such, the relationships between actors are the focus of this analytical tool. Techniques can be applied to isolate individuals that are critical to an intervention. For instance, individuals or groups that have strong:

- **Centrality:** those with many relationships;
- **Prominence:** those with the power and ability to influence networks and individuals; and,
- **Brokerage:** those who can foster entrepreneurial relations or connections between others.⁷⁷

Image 1: A Simple Social Network Analysis



(photo credit: Wambeke, *Social Network Analysis*)

This tool can be used to measure social relationships in conflict and fragile environments. It can show who is connected to whom and the strength of the relationship within the larger network. This tool will help identify who are the most significant actors or organizations that an intervention should target. It can also show which actors or organizations need support to be able to operate more effectively with others. For example, there may be an intervention that seeks to strengthen civil society organizations (CSOs) to more effectively work with government officials to resolve conflict peacefully. The intervention may involve improving the capacity of CSOs and government officials, as well as incentivizing them to work jointly to resolve conflict. Social network analysis may be used to determine the strength and the relationship between CSO and government officials before the programme started and after the intervention. When social network analysis is used during the baseline and the evaluation, it can visually demonstrate changes in complex social relationships over time. This tool may also showcase how an intervention may have contributed to a change in social dynamics.

Typical key questions to ask in your social network analysis include:

- "Who knows who and how well?"
- How well do people know each other's knowledge and skills?
- Who or what gives people information about xyz?
- What resources do people use to find information/feedback/ideas/advice about xyz?
- What resources do people use to share information about xyz?"⁷⁸

⁷⁵ Emirbayer, Mustafa, and Jeff Goodwin. "Network analysis, culture, and the problem of agency." *American journal of sociology*, vol. 99, no. 6, 1995: 1411-1454.

⁷⁶ Emirbayer and Goodwin, "Network analysis," 1411-1454.

⁷⁷ Haythornthwaite, Caroline. "Social network analysis: An approach and technique for the study of information exchange." *Library & Information Science Research* 18, no. 4, 1996: 331.

A common pitfall in social network analysis is the level of knowledge required about the tool to effectively draft the survey and produce the visual with the computer software. There may be a need to raise the capacity of existing staff to utilise this tool. In addition, defining what constitutes a network or system can be time consuming and difficult given complex environments. A snowball sampling methodology may help map the network to make sure that you have a complete network of all the relationships. The protocol must also define the types of relationships that are to be studied and why.⁷⁹

Nevertheless, these challenges can be overcome with strategic and intentional research design. Paired up with a strong stakeholder and conflict analyses, social network analysis could identify key target groups and individuals that interventions should work with to produce the most change.

In summary, social network analysis is a complicated tool for complicated and complex problems. SNA can examine nonlinear social change processes, ripple effects, and unintended, indirect consequences. SNA naturally pairs well with systems analysis (p. 41), stakeholder analysis (p. 31), outcome mapping (p. 32) and most significant change (p. 39). It might even be paired with List Experiments (p. 37), to better understand the relationship between individual characteristics, the networks that an individual is situated in, and larger societal changes.

Stakeholder Analysis

Stakeholder analysis is a simple tool that seeks to (1) identify the key stakeholders and (2) include a wide range of stakeholder perspectives throughout the design, monitoring and evaluation process. It includes the identification of key stakeholders and elicits their knowledge, views, and perspectives on all aspects of an intervention, including their expectations for the resulting changes. Stakeholder analysis may be done utilizing qualitative or quantitative tools.

Conducting the stakeholder analysis can help identify the key individuals, groups, or institutions in a community that are crucial to bring about peace. It can identify the strengths and weakness of the key stakeholders. "In carrying out the analysis, questions are asked about the position, interest, influence, interrelations, networks and other characteristics of stakeholders, with reference to their past, present positions and future potential."⁸⁰

⁷⁸ Wambeke, Tom. "Social Network Analysis." L&T blog: About learning, training and technology at ITC-ILO. <http://itcilo.wordpress.com/2011/03/07/social-network-analysis/> (accessed March 29, 2013).

⁷⁹ Laumann, Edward O., Peter V. Marsden, and David Prensky. "The boundary specification problem in network analysis." *Research methods in social network analysis*, 61, 1989: 87.

⁸⁰ Brugha, Ruairi, and Zsuzsa Varvasovszky. "Stakeholder analysis: a review." *Health policy and planning* 15, no. 3, 2000: 239.

Example 6: Stakeholder Analysis in Ethiopia, Uganda, and Georgia

Stakeholder Analysis is an integral part of Mercy Corps' programming, and has been conducted in a variety of contexts and sectors, including Ethiopia, Indonesia, and Uganda for the Evaluation and Assessment of Poverty and Conflict Interventions (EAPC) Study.

For example, **in Ethiopia** a participatory focus group methodology was utilised to identify the key actors that both mitigated and aggravated conflict and explored the driving factors behind the actor's actions. Triangulated with a series of other participatory tools, Mercy Corps was able to create a robust stakeholder analysis and conflict assessment. Copy of the tools can be found in the following report:

http://www.mercycorps.org/sites/default/files/eapc_lessons_learned_on_measuring_impact_2011-07-28_final.pdf

In Uganda a stakeholder analysis was used in a mobile-based agricultural program that focuses on improving services for small farmers. The stakeholder analysis was used to: ensure sustainability of the program; understand the strengths and weaknesses of each stakeholder; highlight financial incentives to keep stakeholders engaged; and determine appropriate roles and responsibilities within the program. Similarly, Mercy Corps conducted a stakeholder analysis **in Georgia** in 2012 in order to identify problems associated to regional cooperation for climate change adaptation.

(sources: *Understanding Poverty and Conflict*, Mercy Corps; *Evaluation and Assessment of Poverty and Conflict Interventions: Conflict & Economics: Lessons Learned on Measuring Impacts*, Mercy Corps.)

Stakeholder analysis can also help measure the community's perception of key individuals and can be used to see if there is a change after an intervention. In baselines and evaluations, stakeholder analysis helps measure the un-measurable by intentionally including marginalised perspectives and previously unheard voices in order to provide a more nuanced understanding of the key actors involved in the conflict, the changes that an intervention needs to accomplish, and whether these were achieved.

As an evaluative tool, it can uncover the changes the intervention brought about, for whom and with what consequences. This type of analysis will enrich and provide more nuanced analysis, contribution statements, recommendations, and lessons learned.

In Summary, stakeholder analysis is particularly relevant in fragile and conflict-affected environments "where people's sense of security and vulnerability may be heightened and where tensions and factions may exist."⁸¹ The inclusion of stakeholder voices throughout the design, monitoring and evaluation process empowers participants and facilitates the airing and understanding of the 'other's' perspective. For more robust stakeholder analysis, one might integrate the methodology with social network analysis (p. 30) within conflict assessments to more clearly identify key people for the intervention, who they are connected to, and how they relate to one another.

Outcome Mapping

Outcome mapping is a methodology used for design, monitoring and evaluation that seeks to identify changes in attitudes, behaviours, knowledge and perceptions at the outcome level within the direct partners or beneficiaries of a programme. It then analyses the extent of the intervention's *contribution* towards outcome level changes rather than attribution.⁸²

⁸¹ *Handbook on Planning, Monitoring and Evaluating for Development Results*. New York: United Nations Development Programme, 2011.

⁸² Earl, Sarah, Fred Carden, Michael Quinn Patton, and Terry Smutylo. *Outcome Mapping: Building Learning and Reflection into Development Programs*. Ottawa: International Development Resource Centre, 2001.

It is well-suited for the measurement of intangible changes within the immediate influence of a programme, and for learning about how change unfolds in complicated and complex environments. As such, it is appropriate for the development of an evidence base for ‘what works’ at the individual or group intervention level. If integrated across funded programmes, it can illustrate nuances between and across micro-contexts (individual project geographic scopes) and macro-contexts (country or portfolio levels).

One of the outcome mapping’s fundamental principles is that a programme can only *influence* change, not control it, and therefore this approach holds particular promise for adaptive and developmental programming. For example, it can be used to reconstruct or refine programme logic (before or during implementation, or in evaluation) in a participatory manner.

Similarly, outcome mapping facilitates learning and understanding of the system in which the programme occurs, and how change is (or is not) brought about within its immediate sphere of influence. Outcome mapping works best if it is integrated in the monitoring strategies of an intervention. Experience demonstrates that the suggested method for outcome monitoring—journaling—is not widely used and therefore alternative methods may need to be supplemented, including traditional monitoring strategies and tools.

If outcome mapping was not use from the beginning of a programme, an evaluator can use *outcome harvesting*, a related technique in which the evaluator works backward to determine contribution to outcomes by collecting data from reports, interviews, and other sources.⁸³ Since outcome harvesting focuses on what has already been achieved, rather than measuring progress towards predetermined, it is especially applicable in a goal-free evaluations. Outcome harvesting is also useful when there is no baseline data.

Outcome Mapping can help understand value-for-money, effectiveness, and efficiency in long or complex causal chains. For example, an integration of logical framework analysis and outcome mapping tools can help visualise relationships between programme outputs and outcomes, and separate the contributions of the partners towards those outputs and outcomes.⁸⁴ This can be particularly useful in complex causal chains or mechanisms in process evaluation.

⁸³ For a useful guide on outcome harvesting, see: Wilson-Grau, Ricardo, and Heather Britt. *Outcome Harvesting*. Cairo: Ford Foundation, 2012.

⁸⁴ Nyangaga, Julius, and Dr. Edward Ontita. *Outcome Mapping: Reflecting and Learning from applications in eastern Africa*. Nairobi: International Institute of Rural Reconstruction, 2012.

Example 7: Outcome Mapping in East Africa

The International Research Development Centre (IDRC)'s Evaluation Unit funded a study on the use of Outcome Mapping in seven development organizations in East Africa in 2009. This study was used to improve and examine the effects of their Outcome Mapping Manual and local capacity building trainings on the use of this tool for monitoring and evaluation. The Study provides practitioners with real life examples of how outcome mapping can be used.

Case studies in this report offer insights on: how to combine outcome mapping with logical frameworks; translating outcome mapping terminology into more accessible language; responding to unexpected findings; being flexible in journaling and timing; using outcome mapping to enhance relationships with public officials, private organizations, and other stakeholders.

Outcome mapping can be used at all stages of design, monitoring, and evaluation. For instance, the Food and Agricultural Organization (FAO) used outcome mapping to evaluate their efforts to increase disaster and drought resiliency by building local capacity in East Africa. They began by translating the project logic framework into an outcome mapping framework, then collected data using key informant interviews, focus group discussions, a document review, and outcome mapping workshops to assess performance against the resulting outcomes and strategies. OM was used to identify gaps and blockages in programming in order to improve the next phase of the project.

For more information about how to apply outcome mapping, visit: www.outcomemapping.ca

(source: Nyangaga and Onitita, *Outcome Mapping in eastern Africa*, 45-51.)

PART II: SOCIAL SCIENCE METHODOLOGY FOR IMPACT-LEVEL AND LONG TERM MEASUREMENT

Impact Evaluation, Quasi- and Experimental Methods

This section briefly examines three approaches to impact-level evaluation. Much has been written on impact evaluation, quasi- experimental and experimental methods. There is an increasing embrace of these methods in international development, despite numerous literature against and in favour of the methodology. These resources should be consulted and utilised in any effort to use these methods or approaches. Key resources are provided below, but first this section explores the definitions and utility of the approaches and methods.

There are three competing definitions of impact evaluation, broadly speaking. The first, and broadest definition, is from the OECD, whereby an impact evaluation is any evaluation that seeks to assess the “positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended”⁸⁵ within the interventions sphere of influence.⁸⁶ It is essential to note in the operationalization of this definition that impact is relative and contextual, but at minimum is “addressing the macro level factors and dynamics within the intervention’s sphere of influence” (its highest level change[s]).⁸⁷

The World Bank and the Abdul Latif Jameel Poverty Action Lab (J-PAL) disagree by defining impact evaluation according to the *methods* used in the evaluation: quasi-experimental and experimental, respectively.

The World Bank defines impact evaluation as the comparison of “the outcomes of a program against a counterfactual that shows what would have happened to beneficiaries without the programme. Unlike other forms of evaluation, they permit the attribution of observed changes in outcomes to the program being evaluated by following experimental and quasi-experimental designs.”⁸⁸

J-PAL is even more specific, defining impact evaluation as one which use only experimental methods⁸⁹, but recognises that each approach to impact evaluation has its own assumptions, strengths and weaknesses.⁹⁰

Quasi-experimental and experimental methods are approaches to impact evaluation that apply a causal analysis framework to “deliver precise estimates of the cause-effect relationship between policy action and outcomes by comparing predefined treatment and control groups before and after” an intervention.⁹¹ The difference between the two methods is that a quasi-experiment does not use random assignment, whereas experimental methods do, such as in randomised control trials (RCTs). Both tools can be used in complicated and complex measurement problems, particularly for attribution.

⁸⁵ *Glossary of key terms*, OECD-DAC, 24.

⁸⁶ Rogers, Mark M. *Evaluating Impact in Conflict Prevention and Peacebuilding Programs: Working Papers on Program Review & Evaluation*. Cambridge: CDA Collaborative Learning Projects, no. 2, 2012.

⁸⁷ Rogers, *Evaluating Impact*, 2.

⁸⁸ *The Development Impact Evaluation (DIME) Initiative*. The World Bank, http://siteresources.worldbank.org/INTDEVIMPEVAINI/Resources/DIME_project_document-rev.pdf. (accessed March 18, 2013).

⁸⁹ Abdul Latif Jameel Poverty Action Lab. "Impact Evaluation." What is Evaluation?. <http://www.povertyactionlab.org/methodology/what-evaluation/impact-evaluation> (accessed March 18, 2013).

⁹⁰ Abdul Latif Jameel Poverty Action Lab, “Experimental Methodology Table.” <http://www.povertyactionlab.org/sites/default/files/documents/Experimental%20Methodology%20Table.pdf> (accessed March 18, 2013).

⁹¹ Legovini, Arianna. "Development Impact Evaluation Initiative: A World Bank–Wide Strategic Approach to Enhance Development Effectiveness." *Draft Report to the Operational Vice Presidents, World Bank, Washington, DC, 2010*: 5.

As mentioned previously, much has been written on both quasi- and experimental impact evaluation. This paper has no significant further insights to add to these labours. However, the authors recommend the following as particularly relevant for intangible measurement for peace & conflict, security & justice in fragile and conflict-affected environments:

- Stern, Elliot, Nicoletta Stame, John Mayne, Kim Forss, Rick Davies, and Barbara Befani. *Broadening the Range of Designs and Methods for Impact Evaluations: Report of a study commissioned by the Department for International Development*, DFID Working Paper 38, 2012.
- White, Howard and Daniel Philips. *Addressing Attribution of Cause and Effect in Small n Impact Evaluations: Towards an Integrated Framework*, Working Paper 15, International Initiative for Impact Evaluation, 2012.
- Rogers, Mark M. *Evaluating Impact in Conflict Prevention and Peacebuilding Programs: Working Papers on Program Review & Evaluation: #2*, Reflecting on Peace Practice Program, CDA Inc., 2012.
- Shahidur R. Khandker, Gayatri B. Koolwal, and Hussain A. Samad, *Handbook on Impact Evaluation: Quantitative Methods and Practices*, The World Bank, 2010.

While the debate within the peacebuilding and the security & justice sector continues on whether impact evaluation methodologies are appropriate in conflict and fragile environments, several international not-for-profit organizations, such as Search for Common Ground, Mercy Corps, and International Rescue Committee, have all used or plan to use impact evaluations methodologies in their projects in places and circumstances that are appropriate, feasible, and strategic. DFID-funded programmes have also used impact evaluation in such environments.⁹² These efforts should be closely monitored by interested parties, with key lessons learned and insights widely disseminated among the international community.

Natural Experiments

Natural experiments are studies of naturally occurring events in which the outcome of interest was not necessarily planned for (i.e., unintended and unanticipated consequences). Natural experiments are particularly relevant for understanding the ways in which seemingly disconnected factors affect each other and are connected. As such natural experiments are a 'natural' fit for the complex environments and causal chains commonly associated with peace & conflict, security & justice programming.

Natural experiments "come in many forms, including before-after comparisons, cross-section comparisons of treated and untreated groups, or a combination of before-after and group-to-group comparisons",⁹³ and may be, but not necessarily, randomized.⁹⁴ They are mixed methods experiments that may include statistical analysis and traditional tools, such as observation.

Natural experiments cannot be constructed, however; they must be found: "some chance event helps ensure that treatment selection is *not* related to individual characteristics or needs... Thus, in a natural experiment, instead of the usual self-selection or other treatment selection bias that generally occurs, something happens that mimics... a randomized experiment."⁹⁵ For instance, although somewhat disputed as being "as-if" random, researchers and academics have used historical data on population

⁹² *Interim Guidance Note: Measuring and managing for results*, DFID, 26; Stern et al, *Broadening the Range of Designs and Methods for Impact Evaluations: Report of a Study Commissioned by the Department for International Development*.

⁹³ Remler and Ryzin, *Research Methods*, 428.

⁹⁴ Sekhon, Jasjeet S., and Rocio Titiunik. "When Natural Experiments are Neither Natural nor Experiments," *American Political Science Review* 106, no.1, , 2012: 35-37.

⁹⁵ Remler and Ryzin, *Research Methods*, 429.

density, economic development, and export records to serve as part of a natural experiment to evidence the impact of the slave trade on formal and informal institutions.⁹⁶

For example, Christopher Blattman and Jeannie Annan⁹⁷ sought to assess the impact of combat on the human capital of Ugandan youth: the consequences for lifetime labour market performance, and lessons for the economic recovery of civil war-torn countries. The study compared a random comparison group of non-abducted youth to abducted youth in order to understand any human capital differences that might exist between combatants and non-combatants in the post-civil war context of Uganda in 2010. The uniqueness of this case is that the control and treatment groups are relatively well-matched, as the Ugandan rebel forces abducted youth at random, creating an “as-if” randomization for this study.⁹⁸

Natural experiments are particularly relevant when there are practical or ethical constraints to the use of randomized experiments, both of which are objections some practitioners have used against experimental impact evaluations⁹⁹ in conflict settings. Natural experiments are also more amenable to large scale experiments than randomised ones, which enhances the generalizability of the findings.¹⁰⁰ In this way, natural experiments can support the development of systems analysis (see p. 41). They do this by identifying with greater nuance the ways in which variables interact within a system and how they are connected with one another.

A critical factor is that the randomisation occurs from factors external to the study (i.e., naturally occurring in the environment). As such, the identification of existing conditions conducive to a natural experiment for the question at hand can be extremely time-consuming. While natural experiments are able to capture the outcomes of behaviour, they are seldom able to capture what that behaviour in fact consists of, i.e., the contextual meaning of the behaviour; a research question necessitating follow-up qualitative methods such as interviews or surveying.¹⁰¹

List Experiments

List experiments are a quasi-experimental method that aims to elicit accurate aggregate responses of individuals with regards to knowledge, attitudes, behaviours or perceptions that are often considered sensitive or taboo. It is particularly well suited for programming in fragile and conflict-affected environments, or when cultural norms hinder certain tools or lines of inquiry. In this way, list experiments can measure the un-measurable by gathering data on sensitive issues which might not otherwise be available.

The technique is “useful for identifying estimates of the population distribution of responses” but not for identifying individual characteristics that may be associated with that response.¹⁰² This is because the method is meant to offer each participant a degree of anonymity, and the resulting answers are only useful when grouped together.

⁹⁶ Diamond, Jared M., and James A. Robinson. *Natural experiments of history*. Cambridge, Mass.: Belknap Press of Harvard University Press, 2010.

⁹⁷ Blattman, Christopher, and Jeannie Annan. "The consequences of child soldiering." *The review of economics and statistics* 92, no. 4, 2010: 882-898.

⁹⁸ Blattman and Annan, "Consequences of Child Soldiering," 895.

⁹⁹ Experimental impact evaluations are evaluations that normally use randomized control trials.

¹⁰⁰ Remler and Ryzin, *Research Methods*, 428.

¹⁰¹ Paluck, Elizabeth Levy. "The Promising Integration of Qualitative Methods and Field Experiments." *The ANNALS of the American Academy of Political and Social Science* 628, no. 1, 2010: 59-71.

¹⁰² Blair, Graeme. "Methods and Software for Sensitive Survey Questions." Princeton University. <http://graemeblair.com/sensitive/> (accessed March 18, 2013).

List experiments are indirect in their approach to sensitive subjects. This commonly involves creating a randomisation protocol for the assignment of individuals to the control and treatment groups. Respondents are then given a list relating to the subject of investigation, for examples groups and stakeholders in Afghanistan. The respondent is then asked whether s/he agrees with the views of those groups. Respondents do not list the specific groups they agree with, but rather state that out of the five listed groups, they agree with the views of at least two groups. The difference between the control and treatment groups is that an additional stakeholder is added for the treatment group. The added option should be directly related to the purpose of the experiment.

For example, the following scripts might be read to the control and treatment groups in attempting to elicit responses on citizen agreement with competing factions in **Afghanistan**:¹⁰³

Table 4: Example of List Experiment in Afghanistan

Control Group	Treatment Group
I'm going to read you a list with the names of different groups and individuals on it. After I read the entire list, I'd like you to tell me <u>how many</u> of these groups and individuals you broadly support, meaning that you generally agree with the goals and policies of the group or individual. Please don't tell me which ones you generally agree with; only tell me <u>how many</u> groups or individuals you broadly support.	I'm going to read you a list with the names of different groups and individuals on it. After I read the entire list, I'd like you to tell me how many of these groups and individuals you broadly support, meaning that you generally agree with the goals and policies of the group or individual. Please don't tell me which ones you generally agree with; only tell me <u>how many</u> groups or individuals you broadly support.
Karzai Government National Solidarity Program Local Farmers	Karzai Government National Solidarity Program Taliban Local Farmers

Search for Common Ground is using list experiments to measure social norms, attitudes towards gender-based violence, and prevalence of behaviour **in the Democratic Republic of Congo**.¹⁰⁴ Search for Common Ground developed the following script for a questionnaire:

Table 5: List Experiments in the Democratic Republic of the Congo

Control Group	Treatment Group
I'm going to read you a list of things that sometimes happen to people in your community. After I read the entire list, I'd like you to tell me <u>how many happen in your community</u> . Please don't tell me which ones happen, just tell me how many take place.	I'm going to read you a list of things that sometimes happen to people in your community. After I read the entire list, I'd like you to tell me <u>how many happen in your community</u> . Please don't tell me which ones happen, just tell me how many take place.
(1) Young adults get into fights with each other on the streets (2) Parents sometimes hit their children to reprimand them (3) People argue over land	(1) Young adults get into fights with each other on the streets (2) Parents sometimes hit their children to reprimand them (3) People argue over land (4) Husbands sometimes hit their wives

In this way, data can be gathered on public opinion towards the groups, and therefore of their broad policies, tactics, etc. This method may be particularly relevant for assessing mass attitudes to sensitive cultural subjects, such as female circumcision. Paired with a Likert-scale (see p. 15), List experiments may provide a powerful means of assessing cultural shifts.

¹⁰³ Blair, "Methods and Software."

¹⁰⁴ "Evaluation: Fighting against Gender Based Violence in Democratic Republic of Congo" Search for Common Ground. Forthcoming. June 2013.

List experiments have limitations, however. Namely, it provides *estimates* of population views, not direct percentages. Statistical methods have been developed to overcome this flaw that allows for greater investigation of preferences vis-à-vis characteristics of an individual.¹⁰⁵ The second flaw of list experiments is their low level of reliance of the ‘mean responses’ between the two groups to estimate population proportion of answers. Methods and protocols have been developed to allow for greater nuance in the aggregation and analysis of the response differences between the two groups (multivariate regression analysis).¹⁰⁶

Most Significant Change

Most Significant Change (MSC) is a qualitative, participatory monitoring and evaluation methodology that seeks to identify the most significant changes observed in relation to an intervention. MSC can be summed up with the following question: “Looking back over the last month [or other time period], what do you think was the most significant change in.... [particular domain of change]?”¹⁰⁷ It is a tool that can be used for simple and complicated problems, especially for the observation of unintended or unanticipated consequences.

MSC is also applicable in developmental programming, where pre-determined outcomes, and therefore indicators, may not be feasible (complex problems). MSC can be used as both a monitoring and evaluation tool by identifying the changes that are actually occurring and how these relate to the evolving context. It also allows for local valuing of the changes, which may be different from the donor or evaluator’s valuing of the significance, relevance, effectiveness, efficiency, impact, and sustainability of the observed change. The identified changes can then be aggregated overtime and the most significant impact-level changes identified. There are several manuals and guidance notes to help practitioners implement MSC.¹⁰⁸

Global Giving Storytelling Methodology

The Global Giving Storytelling Methodology¹⁰⁹ is a qualitative, participatory monitoring and evaluation approach that places local people at the forefront of measurement efforts as ‘experts.’ It asks people about their context and what types of programming works based on their individual and collective experience. The purpose is to “help local leaders manage the present,” not predict the future.¹¹⁰ It is a simple tool to help manage complicated and complex dynamic by providing real time *feedback loops* to the intervention and other interested parties.

¹⁰⁵ Blair, Graeme, and Kosuke Imai. "Statistical analysis of list experiments." *Political Analysis* 20, no. 1, 2012: 47-77.

¹⁰⁶ Blair and Imai, "Statistical analysis of list experiments," 47-77.

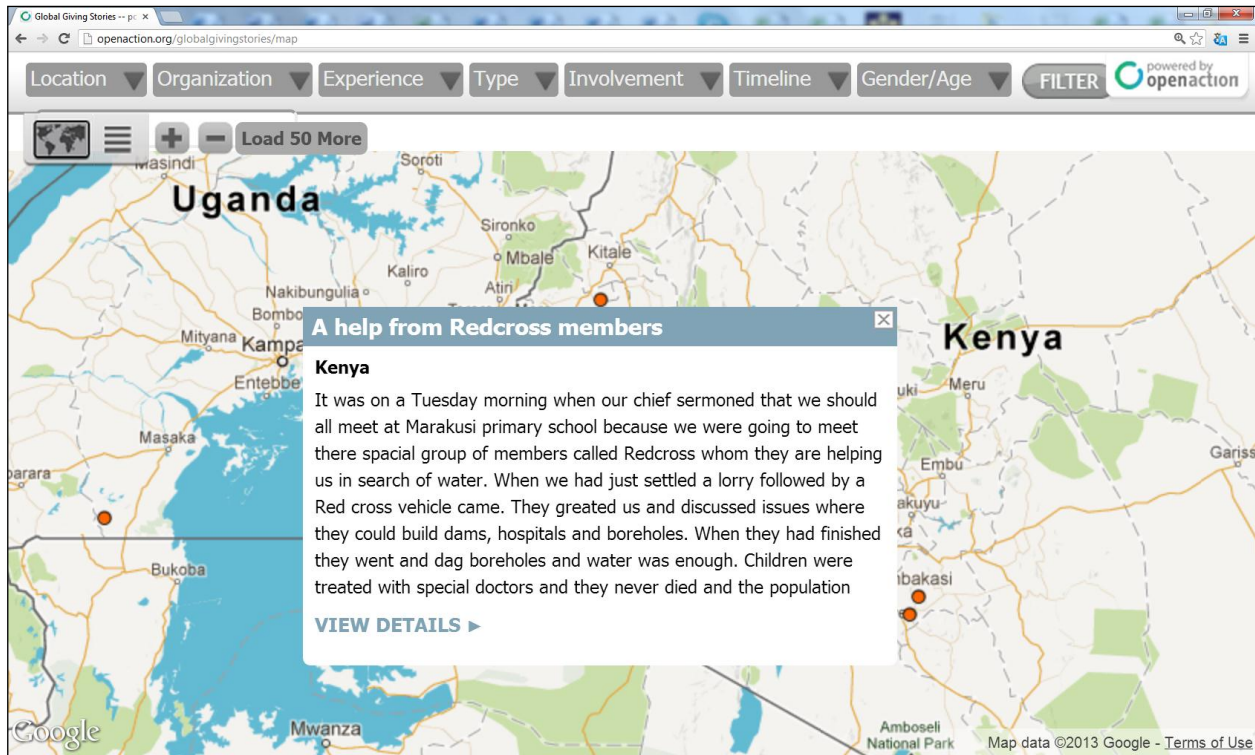
¹⁰⁷ *Strategy Development: Most Significant Change (MSC)*. Overseas Development Institute. London: ODI, 2009.

¹⁰⁸ For further information, please see Davies, Rick, and Jess Dart. *The ‘Most Significant Change’(MSC) Technique: A Guide to its Use*, 2005.

¹⁰⁹ *GlobalGiving Storytelling - Turning Anecdotes into Useful Data*. GlobalGiving: donate to projects around the world supporting disaster relief, education, health, women and children, and more. <http://www.globalgiving.org/stories/> (accessed March 29, 2013).

¹¹⁰ Maxson, Marc, and Irene Guijt. *The “Real Book” for Story Evaluation Methods*. GlobalGiving Foundation and Rockefeller Foundation, 2012.

Example 8: Global Giving Storytelling in Kenya



(source: *GlobalGiving Storytelling - Turning Anecdotes into Useful Data*, GlobalGiving.)

At the heart of storytelling methodologies is the collection of stories from beneficiaries' lives. Stories are collected around structured questions surrounding a particular theme, which may be broad or narrow. Data is input into a computer programme to understand trends, dynamics and results that arise from the aggregation of stories. These results help identify the presence or absence of certain conditions, and dynamics, which can inform the identification of change and determine attribution. Its relatively open-ended nature allows unintended and unanticipated consequences and effects to arise from the identification of patterns across individual stories and geographic scales.

This method has been used to aggregate stories from across wide geographic scales in order to draw meta-conclusions on the types of interventions that work and are the most significant to the individual. Mass application is feasible, as initial analysis suggests this method to be just 5% the cost of traditional data collection methods.¹¹¹ New technologies (see ICTs p. 22) may further support the application of this method at the mass scale. Additionally, the GlobalGiving storytelling method empowers local communities to hold service providers accountable and enhance civic participation.¹¹² Access to Global Giving methodology and manual can be found on their website: <http://www.globalgiving.org/stories/>.

Direct attribution to certain interventions may be possible by comparing storytelling data with other datasets, such as intervention documentation on activities. Those using this method are encouraged to

¹¹¹ Maxson and Guijt, *The "Real Book,"* 4.

¹¹² *GlobalGiving Storytelling - Turning Anecdotes into Useful Data*, GlobalGiving.

open source the data with Global Giving in order to form a more complete picture of ‘what works and under what conditions.’ In this way, it supports the enhancement and use of indices (p. 19) and proxy indicators (p. 20).

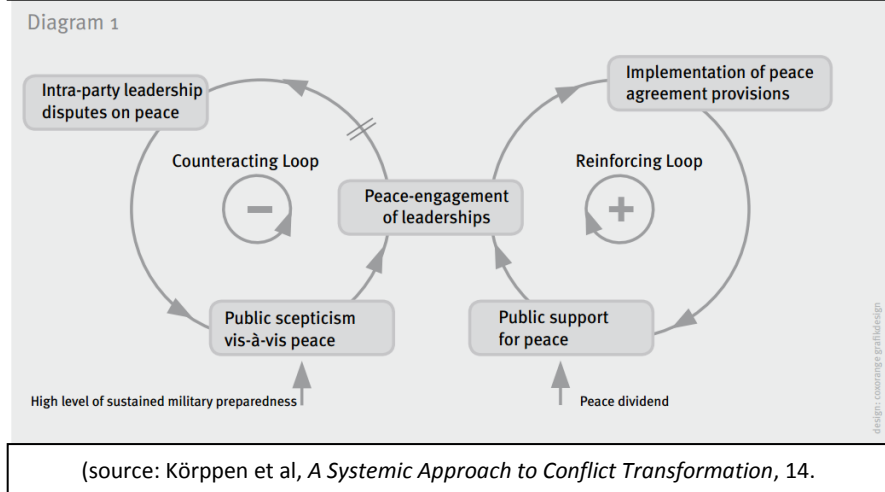
Systems analysis

Systems analysis holds promise for its ability to better understand both the micro and macro-level changes of an intervention within a system. According to Meadows and Wright, “a system is an interconnected set of elements that is coherently organised in a way that achieves something.”¹¹³ For example, the ocean is a complex system that is made up of several organisms that are in constant interaction with one another. The organisms are also interdependent.

A community is another example of a system. It has individuals, groups, institutions that interact with one another bounded by social, cultural, and legal norms. Systems analysis aims to understand the key characteristics of the systems and their interactions. Evaluators use systems analysis to understand the extent to which intervention-based variables (inputs, activities, outcomes) interacted with and affected the overall system and the system components. For example, an evaluator may ask to what extent the capacity building activities of judicial staff and judges have an impact in the overall professionalization of the national judicial system?

There are two basic approaches to systems analysis: mapping and simulations. **Systems mapping** involves identifying the all the key variables within the system, setting their boundaries, and mapping them. The goal is to understand how each variable interact within the system. And plot out the consequences of those interactions on the overall system.

Image 2: Causal Loops Diagram Example



The diagrams consist of two types of causal loops: *reinforcing*, in which the interaction can lead to violence, and *counteracting*, in which the relationship between the variables maintains a healthy system. The result is an identifiable pattern of interactions influencing effects, including their origins and influencers, thus drawing attribution or contribution of intervention variables on system variables.

As illustrated above, the “diagrams help demonstrate and explain the protracted nature of conflict, as well as why peace processes can be ineffective and how to make them more sustainable and transformative.”¹¹⁴

Systems simulations (also referred to as dynamic modelling), on the other hand, uses advanced computer software to simulate the causal model or map of the intervention. The software is capable of

¹¹³ Meadows, Donella H., and Diana Wright. *Thinking in systems: a primer*. White River Junction, Vt.: Chelsea Green Pub., 2008.

¹¹⁴ Körppen, Daniela, Beatrix Schmelzle, and Oliver Wils. "A Systemic Approach to Conflict Transformation." *Berghof Handbook for Conflict Transformation*, Berghof Foundation. 2008.

adding or subtracting variables, and their intensities, in order to arrive at more rigorous conclusions of attribution and effects. It provides a more nuanced perspective on 'success' in that the simulation can provide alternative scenarios. For instance, at times with limited resources one could simulate whether investing in the health-care system or the education system would provide a more stabilizing role in a small community.

In Summary: Both methods help clarify and test our theories and models, identify measurement points, and understand how, why and to what effect nonlinear changes occurred in the system.¹¹⁵ DFID is well placed to employ systems mapping and simulations as a means of devising appropriate funding mechanisms and evaluating country-wide efforts.

¹¹⁵ Kellog, WK, 1.

PART III: SOCIAL SCIENCE METHODOLOGY FOR MEASURING CONTRIBUTION TO PEACE WRIT LARGE

Longitudinal and Cohort Studies

A longitudinal study uses the same data collection tools, engages with the same population, and measures the same things as a regular study, but it is over a long period of time. It is a mixed-methods approach that can pair well with other methodologies, such as focus groups and interviews, as well as impact, quasi-experimental and experimental approaches (see p. 35).

Longitudinal studies are often considered “the best way to assess lasting and sustainable change.”¹¹⁶ They are well-suited to better understand the short-term and long-term social change that peace & conflict and security & justice programmes attempt to produce. It can reveal new insight into the ways in which immediate intervention may have an immediate effect, interact with other variables, and cause unintended or unanticipated consequences in the short-term and long term. This information can be used for statistical analysis or in-depth case studies.

For example, an intervention that aims to change gender norms in a conflict and fragile state may want to use a longitudinal study. After all, gender norms tend to change gradually over time. The M&E team may wish to identify a sample of families and interview on their perceptions, attitudes, and behaviours related to gender norms. The interviewers will return to this sample once a year for the next ten years. The sets of questions will be the same over the period of research. Hence, this longitudinal study will enable the analysis to assess trends in individual changes but also societal changes towards gender norms throughout the ten year period.

Longitudinal studies can continue after the intervention has ended. The continuous monitoring of a certain number of individuals can track changes and long-term impact beyond the life of the programme.

Example 9: Longitudinal Studies in Ethiopia, India, Vietnam and Peru

Young Lives, a collaborative research project primarily funded by DFID and based at the University of Oxford, uses longitudinal and cohort studies, supplemented by surveys, to study childhood poverty in **Ethiopia, India, Vietnam and Peru**. Longitudinal studies track groups of 100 children at 20 different sites in each of these countries in order to document changes in perceptions, relationships, experiences, circumstances, goals, aspirations, and expectations over time. So far, data has been collected in 2007, 2008, and 2011 using a variety of qualitative techniques, such as interviews, community mapping, and daily activity diaries.

According to one of their researchers, "the value of a longitudinal study lies in its potential to explore among the same cohorts of children the impact of home background and early childhood experience on children's educational experience on the one hand, and the impact of educational experience on later outcomes in life such as health, fertility, migration and employment on the other.

(sources: Wilson and Huttly. *Young Lives: A Case Study of Sample Design for Longitudinal Research.*; Crivello, *Longitudinal qualitative research.*)

Its long-term nature and the challenges of gaining consistent access to the same sample make implementing this tool challenging. Often, in conflict and fragile environments populations are

¹¹⁶ Souter, David. "Investigation 4: Impact Assessment." *BCO Impact Assessment Study: The Final Report*. BCO Alliance, 2008: 164.

constantly moving because of the very factors that fuel instability. Sample populations may be refugees or migrant workers. This makes consistent access to the same individuals over time very difficult.

Cohort studies are similar to longitudinal studies, except they occur in shorter periods of time, generally weeks. It is particularly concerned with the individual characteristics (age, religious affiliation, location) that might be associated with particular outcomes (change in attitude or behaviour). It uses a suite of data collection techniques to draw attribution.

Like longitudinal studies, cohort studies often use quasi-experimental methods and other approaches to enhance the rigour of the study. It involves the selection of two groups based on one or more common characteristics (such as ethnicity or association with armed groups) and tracks how these and other characteristics interact with the intervention. For example, it aims to identify whether individuals from one religion are more likely to have a change of attitude than individuals from another religion, if both groups are exposed to the same radio show about conflict resolution. Search for Common Ground, for example, has used cohort studies paired with quasi-experimental methods to attribute changes brought about after exposure to media programming.¹¹⁷

While perhaps the best method for tracking long-term social change, longitudinal studies and cohort studies have not been widely used in international development, and therefore there “is no strongly established evidence base of past experience on which to build.”¹¹⁸ As a result, donors are often hesitant to fund such long term endeavours due to their potential long-term time and monetary commitments. Nevertheless, the experience of other sectors, such as education,¹¹⁹ development work with youth,¹²⁰ which more commonly engage in complex social change, suggest that longitudinal studies hold promise for better understanding the causal mechanisms, sustainability and long-term impact of interventions.

Meta-Analysis

Meta-analysis is a quantitative tool that combines the results of different studies in order to yield new insight into the nuances surrounding outcomes and impacts.

One of meta-analysis’ strengths is its ability to combine results across studies in order to determine typical or standard effect sizes. In other words:

- What is the standard effect (outcomes and impacts, intended or unintended, anticipated or unanticipated, directly or indirectly) that should be expected from a particular type of intervention?
- What do the studies have in common?
- What are the common nuances and connections that can be made across the studies?
- What are the aggregate findings from the studies? What conclusions can be drawn from the patterns that arise in the data?
- What is the likelihood that the change will be sustainable?

“Because meta-analysis summarizes evidence across multiple studies and samples, it produces a better (more accurate, more statistically robust) estimate of the strength and stability of a relationship or

¹¹⁷ See, for example, Abdalla, Amr. *The Team: Kenya: Final Evaluation Report*. Search for Common Ground, 2012.

¹¹⁸ Souter, “Investigation 4: Impact Assessment,” 164.

¹¹⁹ See, for example, Fisch, Shalom M., and Rosemarie T. Truglio. *“G” is for “growing” thirty years of research on children and Sesame Street*. Mahwah, N.J.: Erlbaum, 2000.

¹²⁰ See, for example, Crivello, G. “Longitudinal qualitative research.” *Young Lives: An international study of childhood poverty* <http://www.younglives.org.uk/what-we-do/qualitative-sub-sample-research> (accessed March 21, 2013).

intervention impact than could be obtained in any single study.”¹²¹ By compiling data from more than one study, meta-analysis can overcome traditional critiques of small-n studies and sample sizes. Its application in the fields of conflict & peace and security & justice could be used to establish standard metrics of expected results, which individual interventions can compare results against to make more nuanced and informed judgements on evaluative criteria.

As a statistical method, meta-analysis needs the conversion of qualitative data into quantitative values. If qualitative data cannot be quantified, then it cannot be included in the study. Meta-analysis can also be a time consuming process, requiring a high degree of technical expertise, particularly in quantitative and statistical research methods. A strict coding guide must be developed and presented in such a way that leaves no room for interpretation. Poor coding protocols or errors in the process may threaten the validity of the study.

Meta-analysis is distinct from meta-evaluation. Meta-evaluation can be equated with a peer review process for evaluators to judge the quality of the evaluation “and/or assess the performance of evaluators.”¹²² While the two are distinct, they may overlap. For example, meta-evaluation is frequently used to determine which evaluations are appropriate to include in the meta-analysis.

Conclusion

Measurement and evaluation of the effects of conflict & peace and security & justice interventions is exceedingly difficult. The challenges, while numerous, can be overcome.

This paper outlines some of the key concepts, tools, and approaches from across the social sciences that can help practitioners overcome the challenges of intangible measurement in situations of conflict and fragility. The mixture of tools, methods and approaches can leverage the full weight of the social sciences to overcome the challenges to measuring the un-measurable in situations of conflict and fragility.

Not all measurement challenges are equal, however. Overcoming these challenges first entails clearly identifying what is simple, complicated and complex in the programme theory: how complicated is the causal chain and what within this causal chain is simple or complicated to measure? Knowing these answers allows for the development of appropriate tools based on the purpose and objectives of the measurement.

But, there is no single, catch-all solution to measurement challenges for conflict & peace, security & justice programming. Just as in programme design, each measurement challenge is unique, as is the intended use of that data. Any tool chosen to measure the un-measurable must be appropriate, both for the measurement challenge and the data’s intended use.

Successfully measuring the un-measurable does not rely on using any single tool well, but rather leverages the strengths and weaknesses of a complimentary suite of tools that in combination and culmination lead to measuring of the un-measurable.

¹²¹ Derzon, Jim. “Introduction to Meta-Analysis.” AEA/CDC Summer Institute, 2008.

<http://www.eval.org/SummerInstitute08/08SIHandouts/Uploaded/aea08.si.derzon1.pdf> (accessed 19 March 2013).

¹²² *Glossary of key terms*, OECD-DAC, 27.

Annex A: List of Relevant Indices

- Armed Conflict Location and Event Data (ACLED), <http://www.acleddata.com/>
- Bertelsmann Transformation Index State Weakness Index (Bertelsmann Stiftung/Center for Applied Policy Research, Munich University), <http://www.bti-project.org/home/>
- Center for International Development and Conflict Management International Crises Behavior Version 9, (University of Maryland), <http://www.cidcm.umd.edu/icb/>
- Center for Research on the Epidemiology of Disasters (CRED), <http://www.cred.be/projects>, and the Complex Emergency Database (CE-DAT), <http://www.cedat.be/>
- Center for Systemic Peace, <http://www.systemicpeace.org/>
- Cingranelli-Richards (CIRI) Human Rights Dataset, <http://ciri.binghamton.edu/>
- Country Indicators for Foreign Policy Fragility Index (Carleton University, Norman Paterson School of International Affairs), http://www4.carleton.ca/cifp/app/ffs_data_methodology.php
- Country Policy and Institutional Assessment (CPIA)/International Development Association (IDA) Resource Allocation Index (IRAI) (World Bank), <http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/IDA/0,,contentMDK:21378540~menuPK:2626968~pagePK:51236175~piPK:437394~theSitePK:73154~isCURL:Y,00.html>
- Failed States Index (Fund for Peace), <http://www.fundforpeace.org/global/?q=fsi>
- Freedom House Freedom in the World, <http://www.freedomhouse.org/template.cfm?page=15>
- Global Peace Index (Institute for Economics and Peace, Economist Intelligence Unit, with guidance from an international panel of experts), <http://www.visionofhumanity.org/>
- Guardian Data Journalism and Data Visualization, <http://www.guardian.co.uk/data>
- Harvard Kennedy School Index of African Governance (Harvard University Kennedy School of Government), http://dvn.iq.harvard.edu/dvn/dv/governance/faces/study/StudyPage.xhtml?globalId=hdl:1902.1/13713&studyListingIndex=0_7f27466d655670750bc219f1a2f5
- Hiedelberg Institute on International Conflict Research Conflict Barometer (Hiedelberg Institute on International Conflict Research), <http://hiik.de/en/konfliktbarometer/index.html>
- Human Rights Data Analysis Group, Benetech Human Rights Data Analysis Group (HRDAG), <http://www.hrdag.org/>
- Human Security Gateway, <http://www.humansecuritygateway.info/>
- Index of State Weakness in the Developing World (Brookings Institution/Center for Global Development), http://www.brookings.edu/reports/2008/02_weak_states_index.aspx
- Inter-University Consortium for Political and Social Research, <http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp>
- International Conflict Research Institute (University of Ulster), <http://www.incore.ulst.ac.uk/services/cds/countries/>

- International Institute for Strategic Studies Armed Conflict Database (International Institute for Strategic Studies), <http://www.iiss.org/publications/armed-conflict-database/>
- Lacina/Gleditsch Dataset, <http://www.prio.no/cscw/cross/battledeaths>
- Landmine Monitor, <http://www.lm.icbl.org/>
- National Counterterrorism Center of the US Government Worldwide Incidents Tracking System, <http://www.nctc.gov/>
- Norwegian Peacebuilding Centre FAFO, <http://www.fafo.no/indexenglish.htm>
- Peace and Conflict Instability Ledger (University of Maryland, Center for International Development and Conflict Management), <http://www.cidcm.umd.edu/pc/>
- Peace Research Institute Oslo, <http://www.prio.no/>
 - Center for the Study of Civil War, PRIO, <http://www.prio.no/CSCW/Datasets/>
- Political Instability Index (The Economist Group/Economist Intelligence Unit), <http://www.economist.com/node/13349331>
- Political Terror Scale, <http://www.politicalterrorsscale.org/>
- Project Ploughshares Armed Conflicts Report, <http://www.ploughshares.ca/libraries/ACRText/ACR-TitlePage.html>
- Small Arms Survey (Graduate Institute of International Studies, Geneva, Switzerland), <http://www.smallarmssurvey.org/>
- State Fragility Index (George Mason University, Center for Global Policy), <http://www.systemicpeace.org/polity/polity4.htm>
- Stockholm International Peace Research Institute, <http://www.sipri.org/>
 - SIPRI Arms Transfers Database, <http://www.sipri.org/databases/armstransfers>
 - SIPRI Multinational Peace Operations Database, <http://www.sipri.org/databases/pko>
- The Internal Displacement Monitoring Centre (Norwegian Refugee Council), <http://www.internal-displacement.org/>
- United Nations Resources:
 - United Nations High Commissioner for Refugees (UNHCR), <http://www.unhcr.org/cgi-bin/texis/vtx/home>
 - Peacemaker, <http://peacemaker.un.org/>
- Uppsala Conflict Data Program, http://www.pcr.uu.se/research/ucdp/program_overview/
- US Committee for Refugees and Immigrants (USCRI), <http://www.refugees.org/>
- World Governance Indicators (World Bank Institute), http://info.worldbank.org/governance/wgi/sc_country.asp

Bibliography

Abdalla, Amr. *The Team: Kenya: Final Evaluation Report*. Search for Common Ground, 2012.

Abdi, Dekha Ibrahim . "Peace, Peacebuilding, Peacemaking." In *Berghof handbook for conflict transformation*. Berlin: Berghof Research Center for Constructive Conflict Management, 2001.

Abdul Latif Jameel Poverty Action Lab. "Impact Evaluation." What is Evaluation?.
<http://www.povertyactionlab.org/methodology/what-evaluation/impact-evaluation> (accessed March 18, 2013).

Abdul Latif Jameel Poverty Action Lab, "Experimental Methodology Table."
<http://www.povertyactionlab.org/sites/default/files/documents/Experimental%20Methodology%20Table.pdf>
(accessed March 18, 2013).

Addai, Emmanuel, Emma Kpenu, and Martin Dery. "The Community Scorecard Approach for Performance Assessment: ProNet North's Experience." *WaterAid Ghana*, no. 4, 2004.

Alkire, Sabina. *A Conceptual Framework for Human Security: Working Paper 2*. Oxford: Centre for Research on Inequality, Human Security and Ethnicity, University of Oxford and the Department for International Development, 2003.

A Preliminary Mapping of the Evidence Base for Empowerment and Accountability. London: DFID: Department for International Development, 2011. www.dfid.gov.uk/r4d/PDF/Publications/FINAL_E_and-A_Annex3_Evidence_Mapping.pdf (accessed March 13, 2013).

"ArcGIS - Mapping and Spatial Analysis for Understanding Our World." Esri - GIS Mapping Software, Solutions, Services, Map Apps, and Data. <http://www.esri.com/software/arcgis> (accessed March 21, 2013).

Barnes, Catherine. *Theories of Conflict: Causes, Dynamics and Implications, Applied Conflict Transformation Studies Series*. Birmingham UK: Responding to Conflict, 2005.

Barnett, Chris, Julian Barr, Angela Christie, Belinda Duff and Shaun Hext, *Governance and Conflict Indicators Report*, ITAD and DFID: Department for International Development, 2011.

Bayne, Sarah. *M&E of Integrated Peacebuilding and Development*. DFID: Department for International Development, 2013.

Blair, Graeme. "Methods and Software for Sensitive Survey Questions." Princeton University.
<http://graemeblair.com/sensitive/> (accessed March 18, 2013).

Blair, Graeme, and Kosuke Imai. "Statistical analysis of list experiments." *Political Analysis* 20, no. 1, 2012: 47-77.

Blattman, Christopher, and Jeannie Annan. "The consequences of child soldiering." *The review of economics and statistics* 92, no. 4, 2010: 882-898.

Brugha, Ruairi, and Zsuzsa Varvasovszky. "Stakeholder analysis: a review." *Health policy and planning* 15, no. 3, 2000: 239-246.

Building Peaceful States and Societies: A DFID Practice Paper. London: DFID: Department for International Development, 2010.

Chungong, C., Fortune, F., & Kessinger, A. (2011). Community Radio, Gender & ICTS in West Africa. *Search For Common Ground*, 1, 1-17.

Church, Cheyanne, and Mark Rogers. *Designing for Results: Integrating Monitoring and Evaluation in Conflict Transformation Programs*. Washington, DC: Search for Common Ground, 2006.

Church, Cheyanne A. "Peacebuilding Evaluation: From Infancy to Teenager." *New Routes* 13, 2008.

Corlazzoli, Vanessa, and Jonathon White. *Back to Basics: A Compilation of Best Practices in Design, Monitoring & Evaluation in Fragile and Conflict-affected Environments*. DFID: Department for International Development, 2013.

Corlazzoli, Vanessa, and Jonathon White. *Practical Approaches to Theories of Change in Conflict, Security & Justice Programmes Part II: Using Theories of Change in Monitoring and Evaluation*. 2012.

Crivello, G. "Longitudinal qualitative research." *Young Lives: An international study of childhood poverty* <http://www.younglives.org.uk/what-we-do/qualitative-sub-sample-research> (accessed March 21, 2013).

Davidson, Jane. "Don't expect quantitative evidence to answer a qualitative evaluation question." *Genuine Evaluation*. <http://genuineevaluation.com/dont-expect-quantitative-evidence-to-answer-a-qualitative-evaluation-question/> (accessed March 14, 2013).

Davies, Rick, and Jess Dart. *The 'Most Significant Change' (MSC) Technique.: A Guide to its Use*, 2005.

Day, Melissa, and Joanne Thatcher. "'I'm really embarrassed that you're going to read this...': Reflections on using diaries in qualitative research." *Qualitative Research in Psychology* 6, no. 4, 2009: 249-259.

Derzon, Jim. "Introduction to Meta-Analysis." AEA/CDC Summer Institute, 2008.

<http://www.eval.org/SummerInstitute08/08SIHandouts/Uploaded/aea08.si.derzon1.pdf> (accessed 19 March 2013).

The Development Impact Evaluation (DIME) Initiative. The World Bank, http://siteresources.worldbank.org/INTDEVIMPEVAINI/Resources/DIME_project_document-rev.pdf. (accessed March 18, 2013).

Diamond, Jared M., and James A. Robinson. *Natural experiments of history*. Cambridge, Mass.: Belknap Press of Harvard University Press, 2010.

Diamond, Louise, and John W. McDonald. *Multi-track diplomacy: a systems approach to peace*. 3rd ed. West Hartford: Kumarian Press, 1996.

Earl, Sarah, Fred Carden, Michael Quinn Patton, and Terry Smutylo. *Outcome Mapping: Building Learning and Reflection into Development Programs*. Ottawa: International Development Resource Centre, 2001.

Emirbayer, Mustafa, and Jeff Goodwin. "Network analysis, culture, and the problem of agency." *American journal of sociology*, vol. 99, no. 6, 1995: 1411-1454.

Empowering the Edge: Information Sharing in Post-Civil War Liberia. (2013, January-February). The GIS Professional. Retrieved March 7, 2013, from www.urisa.org/files/janfeb2013.pdf

Evaluating Development Co-Operating Summary of Key Norms and Standards. (2nd ed.). OECD-DAC Network on Development Evaluation, 2010.

Evaluating peacebuilding activities in settings of conflict and fragility improving learning for results. Paris: OECD-DAC, 2012.

Evaluation and Assessment of Poverty and Conflict Interventions: Conflict & Economics: Lessons Learned on Measuring Impact. Mercy Corps.

www.mercycorps.org/sites/default/files/eapc_lessons_learned_on_measuring_impact_2011-07-28_final.pdf (accessed March 13, 2013).

Fisch, Shalom M., and Rosemarie T. Truglio. *"G" is for "growing" thirty years of research on children and Sesame Street*. Mahwah, N.J.: Erlbaum, 2000.

Fragile states 2013: Resource flows and trends in a shifting world. OECD-DAC International Network on Conflict and Fragility (INCAF), 2012.

Freudenberger, Karen Schoonmaker. "Rapid rural appraisal (RRA) and participatory rural appraisal (PRA): a manual for CRS field workers and partners." *Baltimore: Catholic Relief Services*, 1999.

Galtung, Johan. *Theories of Peace: A Synthetic Approach to Peace Thinking*. Oslo: International Peace Research Institute, Oslo, 1967.

Gittelsohn, Joel. *Rapid Assessment Procedures (RAP): ethnographic methods to investigate women's health*. International nutrition foundation (INF), 1998.

GlobalGiving Storytelling - Turning Anecdotes into Useful Data. GlobalGiving: donate to projects around the world supporting disaster relief, education, health, women and children, and more. <http://www.globalgiving.org/stories/> (accessed March 29, 2013).

Glossary of key terms in evaluation and results-based management. OECD-DAC Working Party on Aid Evaluation, 2002.

Glouberman, Sholom, and Brenda Zimmerman. "Complicated and complex systems: what would successful reform of Medicare look like?." *Changing Health Care in Canada: The Romanow Papers 2*, 2002: 21-53.

Goldwyn, Rachel. *Making the Case for Conflict Sensitivity in Security and Justice Sector Reform Programming*. DFID: Department for International Development, 2013.

Goldwyn, Rachel and Diana Chigas, *Monitoring and Evaluating Conflict Sensitivity: Methodological Challenges and Practical Solutions*. DFID: Department for International Development, 2013.

Governance, Security and Justice: Program Prospectus for 2011-2016. DC: International Development Research Centre , 2011.

Greene, Jennifer C., and Valerie J. Caracelli. "Defining and describing the paradigm issue in mixed-method evaluation." *New directions for evaluation* 1997, no. 74, 1997: 5-17.

Handbook on Planning, Monitoring and Evaluating for Development Results. New York: United Nations Development Programme, 2011.

Haythornthwaite, Caroline. "Social network analysis: An approach and technique for the study of information exchange." *Library & Information Science Research* 18, no. 4, 1996: 323-342.

Interim Guidance Note: Measuring and managing for results in fragile and conflict-affected states and situations. London: DFID: Department for International Aid, 2010.

Jaccard, James, and Choi K. Wan. *LISREL approaches to interaction effects in multiple regression*. Vol. 114. Sage Publications, Incorporated, 1996.

Jamieson, Susan. "Likert scales: how to (ab) use them." *Medical education* 38, no. 12, 2004: 1217-1218.

Körppen, Daniela, Beatrix Schmelzle, and Oliver Wils. "A Systemic Approach to Conflict Transformation." *Berghof Handbook for Conflict Transformation, Berghof Foundation*. 2008.

Laumann, Edward O., Peter V. Marsden, and David Prensky. "The boundary specification problem in network analysis." *Research methods in social network analysis*, 61, 1989: 87.

Legovini, Arianna. "Development Impact Evaluation Initiative: A World Bank–Wide Strategic Approach to Enhance Development Effectiveness." *Draft Report to the Operational Vice Presidents, World Bank, Washington, DC*, 2010.

Macnamara, Jim. "Media content analysis: Its uses, benefits and best practice methodology." *Asia Pacific Relations Journal*, vol. 6, no. 1, 2005: 1-34.

Majchrzak, Ann. *Methods for policy research*. Vol. 3. Sage Publications, Incorporated, 1984.

Mateu, Pedro. "Pedro Mateu on Evaluation Meta-analysis · AEA365." AEA365: A Tip-a-Day by and for Evaluators. <http://aea365.org/blog/?p=4138> (accessed March 19, 2013).

Maxson, Marc, and Irene Guijt. *The "Real Book" for Story Evaluation Methods*. GlobalGiving Foundation and Rockefeller Foundation, 2012.

Meadows, Donella H., and Diana Wright. *Thinking in systems: a primer*. White River Junction, Vt.: Chelsea Green Pub., 2008.

Mobile-Based Technology for Monitoring and Evaluation. Clear Initiative Regional Centers for Learning on Evaluation and Results, 2013.

Nyangaga, Julius, and Dr. Edward Ontita. *Outcome Mapping: Reflecting and Learning from applications in eastern Africa*. Nairobi: International Institute of Rural Reconstruction, 2012.

Oakly, Peter, and David Marsden, *Approaches to Participation in Rural Development*. Geneva: International Labour Office, 1987.

Oliphant, Valerie. *Desk Review for Participatory Early Warning to Religious Conflict in Jos Nigeria*. Search for common Ground, 2013.

Paluck, Elizabeth Levy. "The Promising Integration of Qualitative Methods and Field Experiments." *The ANNALS of the American Academy of Political and Social Science* 628, no. 1, 2010: 59-71.

Patnaik, Somani, Emma Brunskill, and William Thies. "Evaluating the accuracy of data collection on mobile phones: A study of forms, SMS, and voice." In *Information and Communication Technologies and Development (ICTD), 2009 International Conference on*, pp. 74-84. IEEE, 2009.

Performance Monitoring & Evaluation TIPS: Selecting Performance Indicators, 2nd ed. Washington, DC: USAID: United States Agency for International Development, no. 6: 2010.

Picard, Mary, and Middle East. "Measurement and methodological challenges to CARE International's rights-based programming." *Enterprise Impact News* 33, 2004.

Reflecting on Peace Practice: Participant Training Manual. Cambridge: CDA Collaborative Learning Projects, 2009.

Remler, Dahlia K., and Gregg Gerard Van Ryzin. *Research Methods in Practice: Strategies for Description and Causation*. New York: Sage Publications, 2011.

Rogers, Mark M. *Evaluating Impact in Conflict Prevention and Peacebuilding Programs: Working Papers on Program Review & Evaluation*. Cambridge: CDA Collaborative Learning Projects, no. 2, 2012.

Rogers, Patricia J. "Using programme theory to evaluate complicated and complex aspects of interventions." *Evaluation* 14, no. 1, 2008: 29-48.

Rogers, Patricia, and Richard Hummelbrunner. "On-line e-Learning programme on: Module 1 Equity-focused Evaluations."

Sartorius, Rolf, and Chris Carver. *Monitoring, Evaluation, and Learning for Fragile States and Peacebuilding Programs: Practical Tools for Improving Program Performance and Results*. Social Impact: 94-97.

Scharbatke-Church, Cheyanne. "Evaluating Peacebuilding: Not Yet All It Could Be." *Berghof Handbook for conflict Transformation, Section II: Analysing Conflict and Assessing Conflict Transformation*, Berghof Foundation: 459-482.

Sekhon, Jasjeet S., and Rocio Titiunik. "When Natural Experiments are Neither Natural nor Experiments," *American Political Science Review* 106, no.1, , 2012: 35-37.

Singh, Janmejay, and Parmesh Shah. "Community Score Card Process: a short note on the general methodology for implementation." *Social Development Department, World Bank*, 2003.

Souter, David. "Investigation 4: Impact Assessment." *BCO Impact Assessment Study: The Final Report*. BCO Alliance, 2008: 164.

Stern, Elliot, Nicoletta Stame, John Mayne, Kim Forss, Rick Davies, and Barbara Befani. "Broadening the Range of Designs and Methods for Impact Evaluations: Report of a Study Commissioned by the Department for International Development." *DFID: Department for International Development*, 2012.

Stone, Christopher. *Supporting Security, Justice, and Development: Lessons from a New Era*. New York: Vera Institute of Justice, 2005.

Strategy Development: Most Significant Change (MSC). Overseas Development Institute. London: ODI, 2009.

Taylor, Sebastian. *Practical Approaches to Strengthening Data and Evaluation in Fragile and Conflict Settings*. theDLgroup, DFID Somalia, 2012.

Understanding Political Violence among Youth: Evidence from Kenya on the links between youth economic independence, social integration, and stability. Mercy Corps, 2011.

Understanding Poverty and Conflict. Mercy Corps. <http://www.mercycorps.org/research-resources/understanding-poverty-and-conflict> (accessed March 29, 2013); Mercy Corps.

United Nations Children's Fund. "Evidence for Children: Developing national capacities for Country-led evaluation systems." United Nations Children's Fund Evaluation Office. www.unicef.org/evaluation/files/DevelopingCapacities_CountryEvaluationSystems_final.pdf (accessed March 20, 2013).

United Nations Development Programme. "Proxy indicator." *Monitoring and Evaluation: Project framework, baselines, indicators and targets*. <http://hrba.undp.sk/index.php/monitoring-and-evaluation/project-framework-baselines-indicators-and-targets/proxy-indicator> (accessed March 21, 2013).

United Nations Evaluation Group. "Evaluation Capacity in the UN System." United Nations Evaluation Group. www.unicef.org/evaluation/files/Evaluation_Capacity_in_the_UN_System.pdf (accessed March 20, 2013).

"Ushahidi." Ushahidi. <http://www.ushahidi.com/> (accessed March 21, 2013).

Violence against Women and Girls: CHASE Guidance Note Series: A DFID Practice Paper. DFID: Department for International Development, no.3, 2012.

Wambeke, Tom. "Social Network Analysis." L&T blog: About learning, training and technology at ITC-ILO. <http://itcilo.wordpress.com/2011/03/07/social-network-analysis/> (accessed March 29, 2013).

Whitehead, Sonia, and Esther Saville. "Using content analysis to measure the influence of media development interventions: Elections training for journalists in Yemen." *BBC World Service Trust: Innovative Research*, 1, no. 01,

2008. http://www.dfid.gov.uk/r4d/PDF/Outputs/MediaBroad/bbcwst_research_series_yemen.pdf (accessed March 14 2013).

Wild, Leni, and Daniel Harris. "More than just 'demand': Malawi's public-service community scorecard." *Overseas Development Institute and Plan International*, no. 69, 2012.

Wilson-Grau, Ricardo, and Heather Britt. *Outcome Harvesting*. Cairo: Ford Foundation, 2012.

Wilson, Ian, and Sharon Huttly. *Young Lives: A Case Study of Sample Design for Longitudinal Research*. London: Young Lives: An international study of childhood poverty, Working Paper no. 10, 2003.

Woodrow, Peter. *Practical Approaches to Theories of Change in Conflict, Security & Justice Programmes Part I*, 2012.

Working Effectively in Conflict-affected and Fragile Situations Briefing Paper I: Monitoring and Evaluation: A DFID Practice Paper. London: DFID: Department for International Development, 2010.