

DESIGNING FOR RESULTS: Integrating Monitoring and Evaluation in Conflict Transformation Programs

Cheyenne Church and Mark M. Rogers



DESIGNING FOR RESULTS: Integrating Monitoring And Evaluation In Conflict Transformation Programs

Cheyenne Church and Mark M. Rogers

© 2006 Search for Common Ground

Duplication and transmission is encouraged and authorized without special permission, as long as acknowledgement is given to Search for Common Ground.

1601 Connecticut Avenue, NW
Suite 200
Washington, DC 20009
Telephone: 1-202-265-4300
www.sfcg.org



The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the United States Institute of Peace or of the Alliance for Peacebuilding.

December 13th, 2005

Dear colleagues in the field of conflict transformation,

It is my pleasure to introduce you to this manual for monitoring and evaluating peacebuilding initiatives. *Designing for Results: Integrating Monitoring & Evaluation in Conflict Transformation Programmes* was produced by Search for Common Ground, an international non-governmental organisation working in the field of conflict transformation, in partnership with the United States Institute of Peace and the Alliance for Peacebuilding. The authors of this manual are Cheyanne Church and Mark M. Rogers, whose hands-on experience, coupled with their deep analytical skills and theoretical bases, have meant that our field now has the first practical manual of its kind to which to turn.

It is my joy to thank both Cheyanne and Mark for the invaluable contribution they are making to our field.

Let me tell you a bit about Search for Common Ground and why we wanted to put this manual together.

Our mission is to transform the way the world deals with conflict: away from adversarial approaches, toward cooperative solutions. Our operating motto is: “Understand the differences; act on the commonalities.”

Since 1982, we have developed comprehensive conflict transformation programmes in: Angola, Burundi, Côte d’Ivoire, the Democratic Republic of Congo, Guinée, Indonesia, Iran, Kosovo, Liberia, Macedonia, the Middle East (with offices in Jerusalem and Amman), Morocco, Nepal, Nigeria, Sierra Leone, Ukraine and the United States. We are also working on the broad issues pertaining to Islamic-Western relations in a large programme called Partners in Humanity.

Today, we have more than 350 full-time staff around the world, which makes us one of the largest NGOs working in our field. Our headquarters are in Brussels and Washington, DC.

We appreciate that people and nations will always act in their perceived best interests, but that everyone’s best interest is served by solutions that maximize the gain of those with a stake in the outcome. Today’s problems – whether ethnic, environmental, or economic – are too complex and interconnected to be settled on an adversarial basis.

We believe that non-governmental organisations like ours can – and should – play a key role in complementing and supplementing the work of governments and multilateral organisations and that close cooperation improves the chances for successful conflict prevention and resolution.

Our core principles include:
Conflict is both normal and resolvable;
Common ground is not the same as compromise;
Conflict can be transformed;
Peace is a process;
Humankind is interdependent.

Our operating practices include:
Cooperative action;
Using an integrated approach;
Committing to engage and discovering the possibilities with time;
Avoiding parachuting;
Being social entrepreneurs;
Being fully immersed in local cultures.

Our operational methods are diverse. Our “toolbox” includes:
Traditional conflict resolution techniques;
Mediation and facilitation;
Capacity strengthening;
Shuttle diplomacy;
Back-channel negotiations;
Practical cooperation projects;
Radio and television production;
Common Ground journalism;
Arts and Culture;
Sports;
Policy Forums;
Polling;
Awards.

As an organisation, we are deeply committed to measuring and increasing the difference our programmes make. We are also committed to being a learning organisation, which shares lessons learned across programmes within the organisation and with our partners in the field, both internationally and locally.

This manual is one of many contributions we hope to make to advance our field. It is our desire that it will prompt wider access to, and use of, the tools and concepts found within. For more information, please visit our website at <http://www.sfcg.org>.

Sincerely,

A handwritten signature in black ink that reads "Sandra D. Melone". The signature is written in a cursive style with a large initial 'S' and a period at the end.

Sandra D. Melone
Executive Director

ABOUT THE AUTHORS

Cheyenne Church is a Visiting Scholar with the Fletcher School and the West Africa Liaison with the Reflecting on Peace Practice project (CDA - Collaborative Learning Project) as an independent consultant. She was previously the Director for Institutional Learning and Research at Search for Common Ground (SFCG). While at SFCG, she spearheaded the use of design, monitoring and evaluation tools as integrated components of peacebuilding programming in SFCG offices from Burundi to Ukraine. She was also involved in a variety of efforts to advance the field of evaluation and peacebuilding, such as the development of a new methodology to measure impact in Macedonia and Kosovo.

Cheyenne has published on evaluation and conflict resolution, single-identity work, and conflict research effects on policy, and most recently, she co-edited *NGOs at the Table: Strategies for Influencing Policy in Areas of Conflict*. She has also taught courses on peacebuilding and evaluation at INCORE (Northern Ireland) and American University (Washington, DC). Prior to SFCG, Cheyenne was the Director of Policy and Evaluation at INCORE. In this period, she was also a member of the Advisory Group for the Reflecting on Peace Practice Project for Collaborative for Development Action. Cheyenne received her MSc from the London School of Economics and her BComm from Queen's University in Kingston, Ontario, Canada.

Mark M. Rogers is an experienced facilitator, trainer, mediator, program designer, and peacebuilder. His main interest is in developing collaborative processes that address community conflicts, strengthen relationships, and end violence. He has taught courses on the practice of peacebuilding at the United Nations University for Peace in Costa Rica and on monitoring and evaluation for peacebuilding at American University. He has also participated in conflict assessments for USAID's Office of Conflict Management and Mitigation and the World Bank.

Previously, Mark served in Burundi as the Country Director for Search for Common Ground's largest and oldest program. Mark has also served as a mediator, trainer, and service coordinator in upstate New York, mediating dozens of neighborhood, family, workplace, and housing disputes and training young mediators for school-based mediation programs.

He has over two decades of field experience in relief and community development in Central and West Africa, Asia, Central America, and the Balkans working with several organizations, including PLAN International, the International Rescue Committee, and the International Medical Corps. Mark holds a Master's in International Administration from the School for International Training in Brattleboro, Vermont and a BA from the University of Colorado.

ACKNOWLEDGEMENTS

Search for Common Ground extends its sincere appreciation to the United States Institute of Peace (USIP) and the Alliance for Peacebuilding for providing funds to support this project.

We are most indebted to Heather Mack, DM&E Project Associate at Search for Common Ground, who managed this project for many months with exemplary patience, attention to details and enthusiasm. This manual wouldn't have been possible without her.

Many individuals gave their time, energy and insights, which have added immeasurably to the quality of this manual. The Readers Group provided important input on the overall structure, and their chapter reviews furthered the authors thinking substantially.

Mohammed Abu-Nimer, Ph.D.

American University, USA

Oscar Bloh

SFCG, Liberia

Diana Chigas

Collaborative for Development Action, USA

Su Flickinger

SFCG, USA

Arijanto Komari

Plan International, Indonesia

Martha McManus

Conflict Transformation Practitioner, Canada

Lena Slachmuis

SFCG, Democratic Republic of Congo

Mary Stewart

DCHA/OTI, USA

Peter Woodrow

Collaborative for Development Action, USA

A dynamic group of 20 Search for Common Ground staff were also critical to the development and articulation of the ideas within the manual. Their patience and constructive criticisms through the manual test process was valued and much appreciated.

Yooke Adelina

Indonesia

Luis Jimbo Kandangongo

Angola

Rebecca Besant

Liberia

Etiole Kazeruke

Burundi

Oscar Bloh

Liberia

Heather Mack

USA

Juanna Bonopha

Sierra Leone

Sandra Melone

USA

Kornelija Cipuseva

Macedonia

Gayle Meyers

Middle East

Annelies Claessen

Belgium

Stephane Mora

Democratic Republic of Congo

Graham Couturier

USA

Nestor Nkurunziza

Democratic Republic of Congo

Aminata Diabaget

Cote d'Ivoire

Annick Nsabimana

Burundi

Abou Fassi-Fihri

Morocco

Alice Rowley

Belgium

Alyona Gorova

Ukraine

Marie Williams

USA

Finally, sincere appreciation needs to be extended to Mana Takasugi whose dedication, intelligence and presence of mind were key to the success of the funding proposal which supported this project. Heather Mack's positive attitude and enthusiasm for the material were not only a pleasure to work with, but she also ensured that the all the research was completed and the details of this project never slipped through the cracks. Thanks also need to be extended to Tim Werner for his eagle eye throughout the editing process.

TABLE OF CONTENTS

Introduction – 1

Chapter 1: Learning – 4

Why is learning important in peacebuilding? – 5

How can we use Kolb's theory of adult learning? – 6

What about combining field and academically based learning? – 8

What about failure? – 9

Chapter 2: Understanding Change – 10

How is success defined? – 12

Who defines what success is? – 13

What are we trying to change? – 13

What are the theories of change for peacebuilding? – 14

How do we use theories of change? – 16

What are the different types of change? – 18

How do we use types of change? – 20

How do theories of change and types of change come together? – 20

Advance Concept: Developing Adaptive Change Processes – 23

Chapter 3: Program Design – 25

Why is design important? – 26

How do we come up with creative peacebuilding program designs? – 27

How are designs built? – 28

How do we set the goal? – 30

How do we manage goals set by the donor? – 31

How do we define the objectives? – 32

How do we select the activities and outputs? – 33

Aren't all design hierarchies relative? – 33

How do we identify assumptions? – 34

How do we put all of these ideas together? – 36

How are logical and results frameworks different? – 37

Advanced Concept: Designing Integrated Peacebuilding Programs – 38

How do we monitor and evaluate if we don't have a design? – 40

If the donors use design terms differently, how do we know what they mean? – 40

Chapter 4: Indicators – 43

What is the purpose of an indicator? – 44

What are the basic components of an indicator? – 45

How do we set the targets of an indicator? – 48

How do we know the indicator will work? – 48

What is the difference between qualitative and quantitative indicators? – 50

What are the other important dimensions for analysis? – 52

What are the risks in working with indicators? – 52

How do we develop indicators? – 53

Advance Concept: Borrowing from Social Capital – 54

If indicators tell us what has changed but not why, how do we find out why it changed? – 57

Advanced Concept: Advanced Indicators – 57

What are the pitfalls to universal conflict transformation indicators? – 59

Chapter 5: Baseline – 61

What is a baseline? – 62

What is the difference between a conflict assessment and a baseline? – 62

How is a baseline utilized? – 64

What does a baseline focus on? – 66

What is a baseline plan? – 68

What does a baseline plan contain? – 69

Who develops the baseline plan and when? – 76

When does a baseline study take place? – 77

Who conducts the baseline? – 78

We have no time to recruit an evaluator for the baseline, what do we do? – 79

What should be done if a complete baseline is not possible? – 80

How do evaluators use baseline data? – 80

Chapter 6: Monitoring – 81

How is monitoring different from evaluation? – 82

Why develop a monitoring practice? – 83

What is context monitoring? – 84

How do we monitor the context? – 84

What is implementation monitoring? – 86

How do we monitor progress towards results? – 87

What do we do with testimonials, anecdotes, and personal narratives? – 88

Why monitor our assumptions? – 89

Should information be collected for each group involved? – 89

How does monitoring fit with logical frameworks and results frameworks? – 89

What do we do when monitoring indicates that we need to make a major programmatic shift? – 90

Example: Restorative Justice for Youth Program Monitoring Plan – 90

Chapter 7: Evaluation Introduction – 92

Chapter 8: Evaluation Preparation — Stage One – 96

Who should be involved in the evaluation preparation? – 98

When does the evaluation preparation occur? – 98

Evaluation Preparation Decision Flowchart – 99

- I. Decision: Evaluation Objectives – 100
 - What do we want to learn? – 100
 - What are the existing criteria or frameworks we can use to guide our thinking? – 100
 - Can the evaluation objectives in the conflict transformation frameworks be explained further? – 101
 - Do we need to use all of the evaluation objectives in the framework? – 105
 - Advanced Concept: DAC Criteria & Conflict Transformation Framework – 105
 - Advanced Concept: RPP 5 Criteria of Effectiveness for Peace writ Large – 107
- II. Decision: Audience – 108
 - Who is the primary audience for the evaluation? – 108
- III. Decision: Formative, Summative or Impact – 110
 - What type of evaluation will it be? – 110
- IV. Decision: Evaluator’s Role – 112
 - What role will the evaluators play? – 112
- V. Decision: Evaluation Approach – 114
 - What is an evaluation approach? – 114
 - Action Evaluation – 114
 - Empowerment Evaluation – 115
 - Goal-Free Evaluation – 116
 - Self-Evaluation – 117
 - Theory-Based Evaluation – 118
 - Utilization-Focused Evaluation – 119
 - What should I consider when selecting the approach? – 122
 - Evaluation Approaches Decision Flowchart – 123
- VI. Decision: Evaluator Scope – 124
 - What is the scope of the evaluation? – 124
- VII. Decision: Evaluator Qualifications – 126
 - Who should conduct the evaluation? – 126
 - Will the evaluators be internal to the organization or external? – 126
 - What type of experience is required? – 128
 - How many evaluators are needed? – 129
 - Will the evaluators be local hires or recruited internationally? – 129
 - Will translation be needed? – 130
- VIII. Decision: Timing – 131
 - When will the evaluation take place? – 131
- IX. Decision: Budget – 132
 - What will the evaluation cost? – 132
 - How long does this process take? – 135

Chapter 9: Evaluation Management – Stage Two – 137

- I. Developing the Terms of Reference - 138
 - What are the terms of reference? – 138
 - When should the term of reference be developed? – 139

Who is involved in developing the terms of reference? – 139

What do the terms of reference contain? – 140

II. The Evaluation Plan - 153

What is an evaluation plan? – 153

Who develops the evaluation plan and when? – 154

What does an evaluation plan contain? – 154

How do you develop the evaluation plan? – 158

Evaluation Plan Example – 159

III. Frequently Asked Questions on Working with External Evaluators - 163

1) When should I start recruiting to get a quality evaluator? – 164

2) Where do I find evaluators? – 164

3) How do I appraise the competence of external evaluators? – 165

4) Is there anything different in a evaluator's contract? – 166

5) What do I need to do before the evaluation begins? – 166

6) How do I know if the evaluation plan is a good one? – 167

7) The security situation has worsened and it is not safe for the evaluation team to travel to some of our worksites. Should we cancel the evaluation? – 168

8) Who should manage the evaluators? – 169

9) How much management is needed for external evaluators? – 170

10) Can I participate in the evaluation of my project? – 170

11) What should I do if I don't agree with the draft report? – 171

12) What do we do with the evaluator's primary data, such as interview notes, after the evaluation is completed? – 172

IV. Strategies for Overcoming Common Evaluation Pitfalls - 172

When the eyes are bigger than the plate – 172

Don't shoot the messenger – 173

Whatever you say, ma'am – 174

Being all things to all people often means being nothing to anyone – 174

We'd be happy to do that for you – 174

We want Wonderwoman/Superman – 175

That logframe is so out of date – 175

It's not a result unless it's a number – 176

I'll just be a fly on the wall – 176

But I thought we were on the same page – 177

Chapter 10: Evaluation Utilization — Stage 3 – 178

When do I start thinking about the use of an evaluation? – 180

What are the steps needed to use content of the evaluation report? – 180

I'm really busy. Can all of this be put together in a checklist? – 185

Advanced Concept: Utilization from a long-term perspective – 187

Chapter 11: Ethics in Evaluation – 188

What is an unethical practice? – 189

What are the ethical challenges common in the design stage? – 190

What are some of the common ethical challenges for baselines and evaluations? – 191
 Protection of People – 191
 Freedom from Political Interferences – 195
 Quality Data-Collection Techniques – 197

What is Informed Consent? – 198

Are there different ethical dilemmas and issues for internal versus external evaluators? – 198

Chapter 12: Methods – 201

What are the basic concepts I need to know about data collection? – 202

 Key Terms – 202

 Quantitative & Qualitative: Methods and Data – 203

What are the standard data collection methods? – 204

How are methods selected? – 210

What do I need to know about instrument development and testing? – 215

What is disaggregated data? – 216

How is data analysis done? – 217

Are there unique peacebuilding tools for data collection? – 218

Can I draw a conclusion for the entire population from this data? – 221

What record maintenance systems are necessary for collected data? – 222

What are ethical obligations of feeding back the results to the people involved? – 222

Conclusion – 224

Appendices

A. Sources for the Terminology Decoder – 227

B. Evaluation Terms of Reference (TOR) Circulation Options – 228

INTRODUCTION

“We are all faced with a series of great opportunities brilliantly disguised as impossible situations.”

- CHUCK SWINDOLL

Much has been learned already...

.....

In Macedonia, our television program targeted children to change their attitudes and behaviors about other ethnic groups in the country. We knew from our monitoring efforts that we had extremely high ratings and that the children both knew the characters and understood the main messages being conveyed. Yet, it was only after our evaluation, which explored the effects of the show beyond the target audience, that we started to understand the true power of the program and the opportunity we had missed to do even more.

The evidence indicated that the television program changed what children considered to be the ideal world in which they wanted to live. It shifted their consciousness about what was possible. Behaviors, however, did not change because children were not able to connect this ideal with the real world in which they lived. Future programs will directly link media work with practical activities for the target audience so that perceptions and behaviors may both change.

.....

Our programming focused on women who had been involved in the resistance movement, and we sought to explore alternative means of catalyzing change. Many of the husbands and brothers of these women were also involved in the resistance. After several months of work with the group, a number of the participants dropped out. At the time, it was deemed normal attrition that one could expect from any type of long-term project. It was only through a formative evaluation that we discovered that these women were being pressured and, in some cases, violently abused at home due to their new ideas and opinions regarding the political situation. Our project had changed them; however, we had not anticipated this consequence. With this information, we were able to design new programs to protect against this unintended negative effect.

.....

One of the objectives of our program was to increase the freedom of movement of the minority community, despite the ongoing state of war. Since the conflict was highly dynamic, we needed real-time information to feed into our programming decisions. By integrating mapping into existing activities, we were able to monitor the changes in people's movements. This helped us understand how perceptions of fear and threat changed over time, which was especially useful because the perceptions of the community were not always the same as ours. We

altered our program based on this information. If we had not done the regular monitoring, we would have missed several important changes.

.....

The authors hope this manual will help peacebuilding practitioners appreciate how design, monitoring, and evaluation (DM&E) can contribute both to their learning and to the success of their programs. The belief that underpins all the concepts within this text is that monitoring and evaluation are the learning disciplines most accessible and most useful for peacebuilding practitioners. Ultimately, excellence in conflict transformation program design and effectiveness is the goal of the authors.

The purpose of the manual is to introduce peacebuilding practitioners to the concepts, tools, and methods needed to incorporate better design, monitoring, and evaluation practices into peacebuilding programming. As an introductory volume, the target audience is front-line peacebuilding practitioners from around the world with minimal formal training in design, monitoring, and evaluation. It assumes the audience has experience, training, and access to resources on conflict assessments, which are a prerequisite to participating in conflict transformation program design.

The belief that underpins all the concepts within this text is that monitoring and evaluation are the learning disciplines most accessible and most useful for peacebuilding practitioners.

A number of factors have contributed to the timing of this manual:

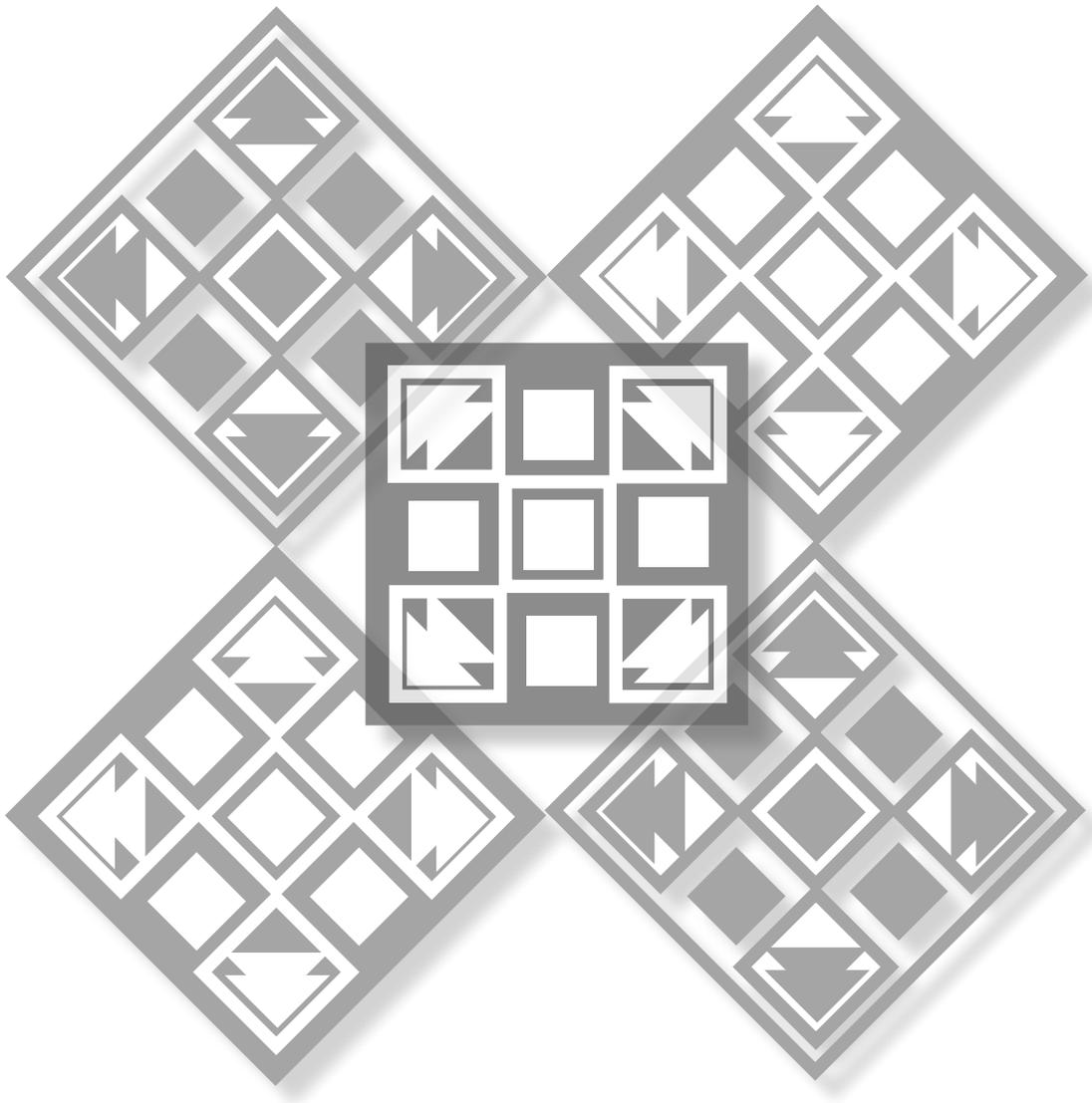
- Myths about the complexity, time, and resources needed to conduct DM&E inhibit programs from seeking out opportunities and building capacity in DM&E.
- DM&E has been approached as an end-game, after-the-fact initiative and, hence, is perceived to be of little value to the practitioner.
- There is a limited pool of seasoned individuals who have a blend of experience in conflict transformation and evaluation expertise.
- The nature of conflict transformation and peacebuilding is qualitative and process-focused, which does not lend itself to quantitative models.
- It is often difficult for peacebuilding practitioners to learn the various DM&E approaches in the detail needed to improve program quality.

The manual offers general information on learning and change in addition to chapters dedicated to specific issues such as baselines, indicators, monitoring, and evaluation. It is organized so that readers can easily jump from one chapter to another. However, we strongly urge jumpy readers to start with the chapter on understanding change because it frames the thinking for most of the discussions in the other chapters.

Every effort was made to offer concrete examples with each of the concepts and methods covered. Many are based upon real programs – predominantly those of Search for Common Ground – while others are fictitious.

The manual is intended to be a living reference, updated periodically.

The manual is intended to be a living reference, updated periodically. We invite readers to share examples of how their DM&E experiences have improved program effectiveness. In finding, adapting, and creating those examples, the authors were again awed at the reach of peacebuilding and its vast array of undertakings and strategies. We hope that readers will be left, as we are, with a thirst for a more in-depth resource since this manual only begins to touch the surface of this complex and exciting field.



LEARNING

This chapter contains:

1. An explanation of the need for learning in peacebuilding
2. The stages of adult learning
3. The advantages of collaboration between scholars and practitioners

INTRODUCTION

“Learning is not attained by chance, it must be sought for with ardor and attended to with diligence.”

- ABIGAIL ADAMS

Learning is “a continuous, dynamic process of investigation where the key elements are experience, knowledge, access and relevance.

Learning is “a continuous, dynamic process of investigation where the key elements are experience, knowledge, access and relevance. It requires a culture of inquiry and investigation rather than one of response and reporting.”¹ This chapter begins with a discussion of the need for learning in peacebuilding. It explains how monitoring and evaluation contribute to learning. It also looks at adult learning, the notion of failure, and the opportunities involved in integrating field-based experience with academic thought.

Why is learning important in peacebuilding?

The work of peacebuilding often requires people involved in a conflict to learn. Peacebuilding program objectives and activities focused on learning abound: raise knowledge, improve understanding, increase tolerance, etc. Changes in relationships require learning new ways to perceive and engage others. Problem solving often involves learning about new options and alternative ways of doing things. The heavy preponderance of training activities is a testament to the fundamental importance of stakeholder learning in peacebuilding.

One of the paradoxes in the field of peacebuilding is that many peacebuilders work so diligently to create learning opportunities for the parties to the conflict, yet they are so casual about their own learning and professional development as peacebuilders.

One of the paradoxes in the field of peacebuilding is that many peacebuilders work so diligently to create learning opportunities for the parties to the conflict, yet they are so casual about their own learning and professional development as peacebuilders. There are several trends that block peacebuilders from learning more about the field and how to become more effective. These include:

- Changes in the conflict context occur quickly and have outpaced our ability to learn from experience.
- We are often too busy doing things right to learn about doing the right things.
- There are too few peacebuilding programs and most of those are too small to effectively rely on informal learning and ad hoc methods of retaining knowledge.
- Rapid strategic changes may result in a loss of knowledge about specific peacebuilding practices.

¹ UNDP Handbook on Monitoring and Evaluation for Results, pp.77.

In addition, peacebuilders use a wide range of excuses to exempt themselves from monitoring and evaluation. Some of the more dog-eared examples include, “Our efforts will only be achieved in the long term” and “With so many actors in the field, it is impossible to isolate our contribution.”

Financial constraints often cause peacebuilding organizations to opt for investing in additional interventions rather than investing in reflection about current programming and the identification of good practices. Obviously, the world needs both additional programming and improvements in programming quality.

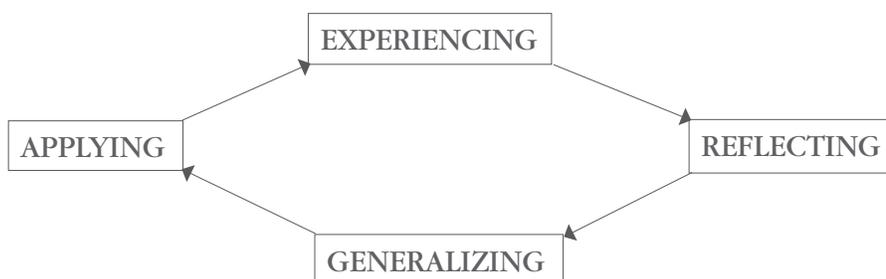
Monitoring and evaluation (M&E) are the learning disciplines most accessible and useful to peacebuilding practitioners. These disciplines are accessible because many donors will pay for them, and they are useful because they are directly related to the practitioners’ experience and context. This manual will provide practitioners with some of the necessary paradigms and tools to undertake M&E, but the ardor and diligence for learning must come from the practitioners themselves.

Monitoring and evaluation (M&E) are the learning disciplines most accessible and useful to peacebuilding practitioners.

How can we use Kolb’s theory of adult learning?²

Most experiential learning activities used in peacebuilding workshops are based on Kolb’s four stages of adult learning: experience, reflect, generalize, and apply. The following illustration of these stages is a systems map rather than a cause-and-effect diagram. The arrows mean “influence” rather than “leads to.” Too often in peacebuilding, practitioners stay firmly planted in the experiencing stage.

Kolb’s Stage of Adult Learning



² For additional thoughts on using Kolb’s stages of adult learning, see *Teaching and Learning: Experiential Learning* by J.S. Atherton (2004). Available at <http://www.learningandteaching.info/learning/experience.htm>.

EXPERIENCE: Adults learn best by both doing and from experience. Yet more experience does not automatically result in better experience or learning. In order to use someone’s experience to full advantage, practitioners need to routinely and systematically move beyond the experience stage in the learning system.

Success merits documentation and so does learning. Anyone who has been through an audit is familiar with the bureaucrat’s refrain, “If it isn’t written down, it didn’t happen.” Undocumented experiences are difficult to share, are more easily forgotten or overlooked, and tend to become exaggerated or distorted over time. Time needs to be allocated to document the how, when, why, and who of an experience so it can be reflected upon later.

“Reflection is the exercise of translating experience into knowledge.”

REFLECTING: While experience may be the best teacher, learning requires more than experience. In processing or reflecting on our experiences, we begin to learn from them. On one occasion or another, most of us have had to experience the same situation several times before we were able to learn enough to either avoid or overcome the situation the next time around. Without processing or reflecting on an experience and learning from it, we are destined to repeat past performance.

Generalizing involves abstract conceptualization.

Monitoring and evaluation can play a critical role in processing our experiences by allowing us to translate experience into information. M&E provides us with the means to reflect on our perceptions and understanding of why things happen and to develop them into increasingly more objective pieces of information and knowledge. Put another way, “reflection is the exercise of translating experience into knowledge.”³

GENERALIZING: Generalizing involves abstract conceptualization. It is a step beyond reflection in that it goes beyond first-hand experience or knowledge of how certain things work to a more general perception about how those things work.

Most peacebuilders derive some sense of satisfaction from their experiences of working with people as they cope with, manage, and transform conflict. The privilege of being present when people risk new means of dealing with conflict comes with its own responsibilities, however. One of those responsibilities is to extract the learning from the experiences and to make it available to peers, other practitioners, and others in conflict.

The other challenge in relying exclusively on experience is that it may reach only small groups of learners, often those who have had the same experience. We do not have the time or the luxury of learning everything

³ International Institute for Rural Reconstruction, *Principles of Community Development*.

through experience, however. This was the premise of the American film “Groundhog Day” in which the protagonist was doomed to repeat the same day endlessly until he had learned everything that that one day had to offer.

Good M&E products enable others to know about and learn from our experiences. As Oliver Wendell Holmes, Jr. once said, “Many ideas grow better when transplanted into another mind than in the one where they sprang up.” Similar data or reinforcing pieces of information enable us to make generalizations about causes and effects, theories of change, spheres of influence, systems, relationships, and attributions. Such information also helps us to understand how conflict, collaboration, peace, equity, and justice interact and function.

APPLYING: Applying new learning and knowledge allows us to modify old behaviors and practice new behaviors in everyday situations – potentially generating new experiences and nourishing the learning system. For additional information on applying learning, see page 178 on Evaluation Use in this manual.

What about combining field- and academically-based learning?

Common arrangements involve partnerships between field practitioners and university staff members and/or researchers. Professors and researchers can offer field practitioners the discipline and tools needed to ensure learning. The field practitioners can provide professors and researchers with data and insights into field realities, constraints, and opportunities. University collaborators, researchers, and professional evaluators can add value to practitioners’ learning in a number of different ways including:

- Helping to ensure adequate standards of rigor in research
- Assisting in selecting and refining the research questions
- Facilitating the analysis of data
- Developing good baseline studies
- Documenting experiences (i.e., process documentation or case study preparation)
- Conducting literature research
- Mapping logical thinking, and
- Lending credibility.

Too many peacebuilding programs fail to make changes, enrich learning, or both.

... peacebuilders, like people in conflict, need opportunities to learn about their work.

What about failure?

Too many peacebuilding programs fail to make changes, enrich learning, or both. Successful and unsuccessful peacebuilding programs that do not translate their experiences into knowledge miss an important opportunity to help others learn from that experience. Unsuccessful programs that generate and disseminate important insights, new tools, or a richer understanding of the dynamics at play can make substantial contributions to the field of peacebuilding and to local peacebuilding practice despite the demise of those programs.

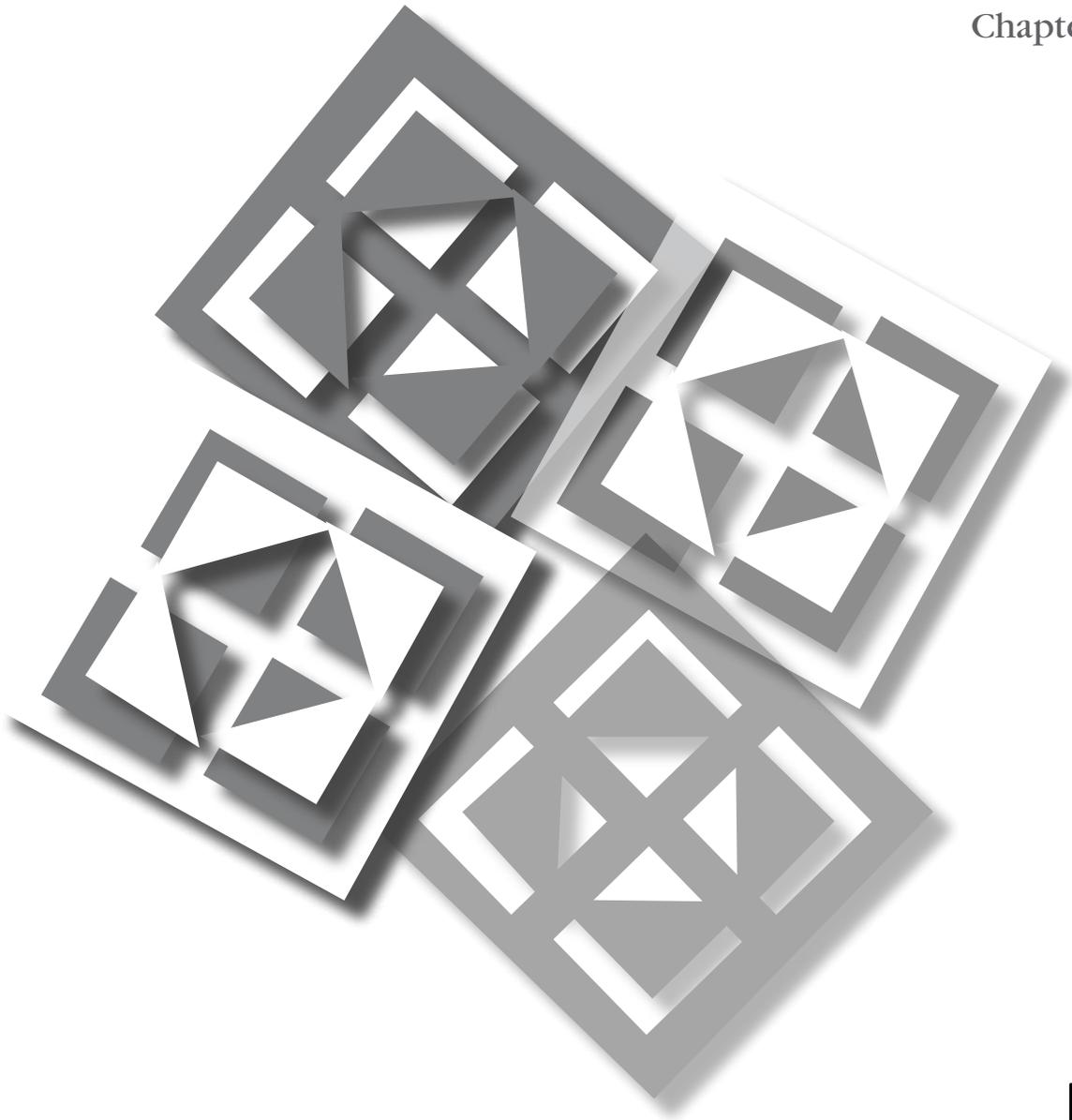
Thomas Edison noted that “many of life’s failures are people who did not realize how close they were to success when they gave up.” Monitoring and evaluation can help practitioners know when to give up and abandon ineffective programs and when to invest energy and resources into initiatives that truly will help.

In sum, peacebuilders, like people in conflict, need opportunities to learn about their work. Monitoring and evaluation are learning disciplines that can contribute a richer understanding of the dynamics at play, greater professionalism, and more effective programming. Collaboration between scholars and practitioners can be mutually enriching.

Further reading

David A Kolb, *Kolb’s Theory of Adult Learning*, 1984.

<http://www.learningandteaching.info/learning/experience.htm>



UNDERSTANDING CHANGE

The chapter includes:

1. Thoughts on defining success
2. Descriptions of theories of change
3. Descriptions of types of change
4. Examples of outcomes

INTRODUCTION

“It is in changing that things find purpose.”

- HERACLITES

A theory of change is a set of beliefs about how change happens.

Answers to what makes peacebuilding successful have been slow in coming. In part, this is due to the intangible dimensions of peacebuilding such as relationships, emotions, communications, identity, values, and culture. Add to this the context-specific idiosyncrasies of each conflict and it is no wonder practitioners often view peacebuilding more as an art and outsiders view it as almost mystical (or just dumb luck). Despite the lack of answers, the practices of the peacebuilding field are considered less a science than those in other fields only because many of us who are practitioners have been less than scientific in our work, studies, learning, and reflection.

This handbook does not offer universal instructions about what to change in building peace. Instead, it offers a number of paradigms to help in our thinking and speaking about design, monitoring, and evaluation of peacebuilding. Behind every peacebuilding initiative there is at least one theory of change. A theory of change is a set of beliefs about how change happens. For example, some believe that culture changes when a critical mass of people takes on new values or morals. Often the theories of change remain implicit, unstated, and unexplored. It is possible to undergo a fairly complete program design process, including goals and objectives, and never examine the underlying assumptions about how change really happens in a given context.

Ideally, practitioners should perform a thorough analysis of the context of a conflict and determine, in consultation with multiple local and international actors, what actions are likely to produce changes in the conflict system. In this context, for instance, will political actors be susceptible to international pressures, economic demand, public opinion, or some other intervention? What drives the decision making by those with the power to make decisions for or against peace?

Too often we are driven in our program choices by our favorite methods – training, dialogue, trauma healing, political negotiation, grassroots mobilization – without considering which of these has the greatest likelihood of leveraging the desired change in the conflict. Program effectiveness is tied to a clear understanding of the ways that change happens in the particular context.

There are no shortcuts and no substitutes for thorough and thoughtful conflict assessment and analysis. The depth and focus of the analysis influences the choice of what to change and what types of change are needed. Designing peacebuilding programs and projects without a complete conflict analysis is irresponsible and potentially dangerous. Before going further, dig up, dust off, update, or complete an assessment using the methodology that best fits your skills, resources, and preferences as well as the conflict on which you are working. Conflict assessments and analysis are not done in a vacuum, however. Be aware of the authors' bias and predisposition – including your own.

An ever-increasing number of conflict analysis tools is available for practitioners to use. Descriptions of many of the more common methodologies can be found in *Conflict-Sensitive Approaches to Development, Humanitarian Assistance and Peacebuilding*, published by Safeworld and partners. The best and most thorough source of information about these methodologies is on the website of the sponsoring organization. A number of the more frequently used models include:

- Do No Harm – Collaborative Development Associates
- FAST methodology – Swiss Peace
- Strategic Conflict Assessment – DFID
- Conflict Analysis Framework – World Bank
- Conflict Assessment Framework – CMM USAID
- Conflict and Policy Assessment Framework – Clingendael
- Conflict Analysis and Response Definition – FEWER

How is success defined?

Like development, peacebuilding strives to make, maintain, or prevent change. This implies some sort of continuum from one point to another. In a dialogue, for example, one party's initial understanding of the dynamics involved in the conflict may be limited to their own positions. As the dialogue continues, they are able to articulate their interests and the others' positions. Further into the process they may be able to express the legitimacy of the others' interests. Here the change sought is one of appreciation of the others' interests. Success is an arbitrary determination of progress and can be set at any point along the continuum in the desired direction of change.

Determining success first requires us to identify what changes are needed, which requires asking, what is the continuum and where are the stakeholders on the continuum? Only after answering these questions can we say how much change needs to take place to be considered a success.

However one chooses to define success, this manual is geared toward looking for success at the level of objectives and outcomes, rather than

There are no shortcuts and no substitutes for thorough and thoughtful conflict assessment and analysis.

Success is an arbitrary determination of progress and can be set at any point along the continuum in the desired direction of change.

Ultimately, the definition of success is the responsibility of the people in conflict.

goals and impact. Information about outcomes can more easily be obtained, tracked, and readily used by practitioners. Measuring impact usually requires sophisticated data collection and analysis methods from multiple sources over extended periods of time. Invariably these requirements either exceed the capacity of many organizations practicing peacebuilding or they extend beyond the donors' funding period. Such limitations explain the importance of partnering with universities and expert researchers who are willing to track, measure, and document the progress of multiple practitioners over longer periods of time.

Who defines what success is?

Ultimately, the definition of success is the responsibility of the people in conflict. Unilaterally, parties will often define success as the total domination of their position. Even though the negotiation may have been conducted by track one or elite groups, the general population eventually accepts or rejects the proposed solutions. For many peacebuilding and conflict resolution processes, a mutually agreed-upon definition of success or goal jointly set by the parties is an important process milestone in and of itself.

In the absence of a consensus among stakeholders about what success would be for a peacebuilding program, donors and NGOs often forge their own definitions. In the best of situations, the discussion about success involves listening to the parties, creating greater opportunities for engagement and participation, and keeping definitions of success broad enough for all stakeholders to easily see how their interests are addressed.

What are we trying to change?

This chapter covers two perspectives on change: theories of change and specific types of changes. Theories of change help planners and evaluators stay aware of the assumptions behind their choices, verify that the activities and objectives are logically aligned, and identify opportunities for integrated programming to spark synergies and leverage greater results. Types of change refer to specific changes expressed in the actual program design and/or evaluation, either as goals, objectives, or indicators. Common examples include changes in behavior, practice, process, status, etc. Both the theory of change and the types of changes sought should be evident in a well-designed program.

What are the theories of change for peacebuilding?

For our purposes, we are using the theories of change for peacebuilding that were developed as part of a large, multiparty action-reflection process called Reflecting on the Practice of Peace or RPP. Among RPP's many participants, there was a consensus that all theories of change are important and necessary; however, different theories may yield greater results under different circumstances.

THE INDIVIDUAL CHANGE THEORY: Peace comes through transformative change of a critical mass of individuals, their consciousness, attitudes, behaviors, and skills. [Methods: investment in individual change through training, personal transformation/consciousness-raising workshops or processes; dialogues and encounter groups; trauma healing.]

THE HEALTHY RELATIONSHIPS AND CONNECTIONS THEORY: Peace emerges out of a process of breaking down isolation, polarization, division, prejudice and stereotypes between/among groups. Strong relationships are a necessary ingredient for peacebuilding. [Methods: processes of inter-group dialogue; networking; relationship-building processes; joint efforts and practical programs on substantive problems.]

THE WITHDRAWAL OF THE RESOURCES FOR WAR THEORY: Wars require vast amounts of material (weapons, supplies, transport, etc.) and human capital. If we can interrupt the supply of people and goods to the war-making system, it will collapse and peace will break out. [Methods: anti-war campaigns to cut off funds/national budgets; conscientious objection and/or resistance to military service; international arms control; arms (and other) embargoes and boycotts.]

THE REDUCTION OF VIOLENCE THEORY: Peace will result as we reduce the levels of violence perpetrated by combatants or their representatives. [Methods: cease-fires, creation of zones of peace, withdrawal/retreat from direct engagement, introduction of peacekeeping forces/interposition, observation missions, accompaniment efforts, promotion of nonviolent methods for achieving political/social/economic ends.]

THE ROOT CAUSES/JUSTICE THEORY: We can achieve peace by addressing the underlying issues of injustice, oppression/exploitation, threats to identity and security, and people's sense of injury/victimization. [Methods: long-term campaigns for social and structural change, truth and reconciliation; changes in social institutions, laws, regulations, and economic systems.]

THE INSTITUTIONAL DEVELOPMENT THEORY: Peace is secured by establishing stable/reliable social institutions that guarantee democracy, equity, justice, and fair allocation of resources. [Methods: new constitutional and governance arrangements/entities; development of human rights, rule of law, anti-corruption; establishment of democratic/equitable economic structures; economic development; democratization.]

THE POLITICAL ELITES THEORY: Peace comes when it is in the interest of political (and other) leaders to take the necessary steps. Peacebuilding efforts must change the political calculus of key leaders and groups. [Methods: raise the costs and reduce the benefits for political elites of continuing war while increasing the incentives for peace; engage active and influential constituencies in favor of peace; withdraw international support/funding for warring parties.]

THE GRASSROOTS MOBILIZATION THEORY: “When the people lead, the leaders will follow.” If we mobilize enough opposition to war, political leaders will have to pay attention. [Methods: mobilize grassroots groups to either oppose war or to advocate for positive action; nonviolent direct action campaigns; use of the media; education/mobilization efforts; organize advocacy groups; dramatic events to raise consciousness.]

THE ECONOMICS THEORY: As a politician once said, “It’s the economy, stupid!” People make personal decisions and decision makers make policy decisions based on a system of rewards/incentives and punishments/sanctions that are essentially economic in nature. If we can change the economies associated with war making, we can bring peace. [Methods: use of government or financial institutions to change supply and demand dynamics; control incentive and reward systems; boycotts.]

THE PUBLIC ATTITUDES THEORY: War and violence are partly motivated by prejudice, misperceptions, and intolerance of difference. We can promote peace by using the media (television and radio) to change public attitudes and build greater tolerance in society. [Methods: TV and radio programs that promote tolerance; modeling tolerant behavior; symbolic acts of solidarity/unity; dialogues among groups in conflict—with subsequent publicity.]⁴

ADD YOUR OWN: This list is in no way comprehensive. Many initiatives have their own theory of change. What is important is to be able to articulate the thinking about how change happens. It need not fit into any of the above theories.

⁴ Peter Woodrow, *Strategic Analysis for Peacebuilding Programs* (excerpt from a longer paper in draft).

How do we use theories of change?

a) To reveal and understand assumptions

Two assumptions are inherent in each theory: 1) how change works, and 2) the strategic advantage of the chosen theory over other theories for the context. Change on the level of the political elites assumes that they remain in power to ensure the advancement of the peace process. It also assumes that the change will endure throughout transitions of power (even after they have left power). Reducing the resources for war assumes that, without weapons, people are less likely to use violence, or that the violence used will result in less death and destruction.

Institutional development, for example, is relatively slow and is often discarded in favor of more immediate results focused on the reducing the level of violence. Where the peacebuilding organization has good relationships with all parties in a conflict, it may give priority to those healthy relations rather than working to reduce the resources for war. The assumption is that, given the analysis, skills, processes, and other assets the organization brings to the situation, it can have the greatest influence using one or two particular theories of change.

b) To ensure alignment in all levels of the program design

When setting goals and objectives, consider the theory of change behind your choices. Are there other theories of change that are better suited to the situation on which you are working? How will effectiveness increase if multiple theories are integrated into the design?

The theory(ies) of change should be discernible at all levels of the program design. If the overall goal of the program is institutional development that targets the court system, for example, then activities directed at reducing the resource for war would seem out of place. One of the common threads that aligns activities, objectives, and goals is the theory of change. Discussions of the underlying theories of change can help tighten program logic and identify gaps and unmet needs.

Certain types of program interventions fit neatly within one theory of change. The example below helps to illustrate the variety of interventions that may fit within any given theory of change. This example also illustrates the overlap between theories of change. For instance, security reform also fits under the institutional development theory as well as the reduce resources for war theory.

Examples of Interventions within Two Different Theories of Change

| | Theory of Change: Reduce the Resources for War | Theory of Change: Public Attitudes Theory |
|----------------------------|---|--|
| HUMAN RESOURCES | Peacekeeping Security Reform Demobilization of combatants Social reintegration of ex-combatants Strengthen resistance/protection of groups vulnerable to violence (Skill building of youth) Disobedience (Israeli military refuseniks) | Train journalists Train influential people and other salient referral sources Involve celebrities and cultural icons Media Literacy |
| MATERIAL RESOURCES | Small Arms and light weapons reduction, control, registration, etc. Disarmament Demilitarized zones (Korea) | Radio and TV programming Social marketing campaigns Cultural, social and sports gatherings Distance learning resources Textbooks and curricula |
| FINANCIAL RESOURCES | Limit and restrict resources that can be diverted into war resources (Congo's natural resources) Transparency in trade (Blood diamonds in West Africa) | Purchase air time and other communication channels Invest in media plurality |

The above examples are neither exhaustive nor restrictive. They are offered simply to inspire program designers to consider the wide range of options available among theories of change. The use of human, financial, and material resources as a means to organize different interventions was borrowed from common practices in facilitating participatory evaluation of community development projects

c) To promote coordination and integrated programming

Giving emphasis or priority to one theory of change does not imply that the others are without merit. More often, the choice represents the capacity or principles of the peacebuilding organization. Peacebuilding requires numerous initiatives at many levels. Integrating different initiatives from multiple organizations that are using several theories of change would seem to offer the best prospects for peace. In this context, integration means the coordination and synthesis of multiple peacebuilding programs, rather than the integration of peacebuilding programs with development or relief programs. More on integration of programming is found in the chapter on Design page 25.

Types of change refer to categories of that can be altered (e.g., changes in knowledge, changes in behavior). Programs often sequence different types of change. For example, a relationship-building initiative might consider knowledge about history to be a prerequisite to changes in stereotyping behavior. The types of change are not totally distinct; some overlap. The intent is not to create definitive academic distinctions, but rather to inspire creative, thorough, and strategic design.

Types of change refer to categories that can be altered (e.g., changes in knowledge, changes in behavior).

The following table illustrates some of the many examples of specific changes that comprise each of the major types of change.

What are the different types of change?

Well-done conflict assessments and analyses are instrumental in identifying the types of change needed in a given context. While the theories of change refer to the broad strategies behind different approaches to peacebuilding, almost all peacebuilding programs implicitly target specific changes in people, communities, organizations, institutions, cultures and/or societies. These more-specific changes help in articulating goals and objectives and in developing indicators for monitoring and evaluation.

Well-done conflict assessments and analyses are instrumental in identifying the types of change needed in a given context.

Examples of Types of Change⁵

| Type of change | Examples of specific changes |
|---------------------|--|
| Relationship | From adversaries to partners in problem solving From suspicion to solidarity From different ethnicities to a common nationality Former neighbors reconciled |
| Status | Soldier to veteran From rebel leader to parliamentarian From entrepreneur to criminal |
| Behavior | From violent behavior to assertiveness From disrespecting women to respecting women From ignoring youth to taking their interests into consideration |
| Circumstance | From politically marginalized to able to vote From displaced hurricane victim to community member |
| Functioning | Increasing transparency From authoritarian to consultative policy development Increasing cost efficiency |

⁵ Modified from *Outcome Examples*, Michael Quinn Patton, *Utilization-Focused Evaluation*, 3rd Edition, Sage, 1997.

| Type of change | Examples of specific changes |
|---------------------|--|
| Attitude | Greater tolerance of different perspectives From fear of others to trust in others From apathy and fatalism to hope and self-determination From a narrow focus on the neighborhood to a broad focus on inter-communal interests |
| Knowledge | Understands interdependence of groups Understands how globalization affects local livelihoods Understands rights and how justice systems should work Knows how political resources are allocated |
| Skills | From power-based to interest-based negotiations Moving discussions from mutually exclusive interests to framing issues in mutually acceptable language Able to introduce items onto the agenda in local governance |
| Maintenance | Continue to celebrate cultural heritage Maintain existing social cohesion Continue to practice traditional dispute resolution processes |
| Prevention | Peaceful transfer of power Increase awareness of military accountability to civilian ministries Prevent exodus of trained and educated professionals |
| Process | From shuttle diplomacy to face-to-face negotiation From hate-mongering to balanced reporting From divisive methods to methods that bring people together From concentrations of authority over others to equitable engagement with others |
| Structural | Creation of a Ministry of Peace New office of Alternative Dispute Resolution established in Ministry of Justice |
| Add your own | |

How do we use types of change?

For example, consider a program aimed at reducing the flow of small arms and light weapons across Kashmir’s borders. The theory of change is “withdraw the resources for war.” The assessment indicates that customs officials along the borders turn a blind eye to illicit arms flows. One of the objectives of the project is to support customs officers to more stringently apply customs regulations relating to arms shipments – a change in behavior. Project planners anticipate seeing an increase in seizures of contraband arms.

These more-specific changes help in articulating goals and objectives and in developing indicators for monitoring and evaluation.

North and South Korea offer another example of how to use types of change. In this case consider a program aimed at increasing the level of international collaboration around issues of mutual interest such as energy, fishing, or trade with Asian neighbors. The theory of change is the Health Relationship/Connections Theory. The specific change, however, focuses on changes in the way international relations function, from antagonistic mud-slinging to negotiated processes that advance mutual interests. The type of change is functioning. Project managers hope to see an increase in the number of bi-lateral memoranda of understanding.

How do theories of change and types of change come together?

In order to develop programming options, program designers take into account the conflict assessment, their explicit theories of change, and the specific types of change desired. The attached table provides an example of each type of change for each theory of change. It is easiest to read the table by starting at the top and reading down along each theory of change (down the column) completely before moving to the next column.

For example, consider the unrest in France in the autumn of 2005. If the assessment indicated that the conflict was primarily driven by exclusion and discrimination and the theory of change was Healthy Relationship, what type of change might be strategic? The table offers only a few examples of literally thousands of possibilities. The intent is not to provide you with the definitive answers, but to stimulate your thinking and your consideration of a vast array of possibilities. The wide range of alternatives is part of the reason peacebuilding initiatives are so difficult to design and to evaluate.

Examples of Potential Outcomes within each Peacebuilding Theory of Change

| Type of Change | Reduce Resources For War | Work with Elites | Mobilize Grassroots | Healthy Relationships | Address Root Causes | Reduce Violence | Individual Change | Institutional Change |
|-------------------------------|---|---|---|---|--|--|---|--|
| <i>Change in Circumstance</i> | Fewer small arms and light weapons enter country | Increased opportunity for popular participation in local governance | Oppressed groups able to voice interests in media | Segregation barriers removed | All groups able to put issues on public policy agenda | Increase security to allow greater mobility | Able to circulate in previously restricted areas | Reduced allocation of national budget to the military and defense |
| <i>Change in Status</i> | Illegal arms importers recognized as criminals | From inherited status to merit-based status | Oppressed groups shift from objects to subjects of change | Relations built on interests rather than status | Elimination of the criteria for exclusion | Increase opportunities to become a part of a non-violent social group | Able to access previously restricted services | Civilian truth & reconciliation commissioners serve in Ministry of Justice |
| <i>Change in Behavior</i> | Customs officials seize illegal arms imports | Elites allow greater participation of subordinates | Increase in non-compliance with oppressive controls | Increase in joint social activities | Increase focus on the issues and less criticism of individuals | Increase use of third party neutrals to resolve potentially violent disputes | Will comply with laws and regulations | Promotions based on merit |
| <i>Change in Functioning</i> | Improvement in regulations affecting small arms and light weapons | Increase in transparency | Agenda-setting open to all | Routinely exchange information | Increased independence of the judiciary | Increasing employment of youth vulnerable to violence | Able to act as a neighbor rather than an opponent | Political appointments no longer the exclusive domain of the executives |
| <i>Change in Structure</i> | Customs bureau separated from taxation bureau | Admissions to higher education no longer exclusively for elites | Free primary education for all | Equitable, affordable access to health care | Separation of judiciary and executive branches | Reduce structural violence stemming from food insecurity | Understand dynamics of structural violence | Truth & reconciliation commission able to grant amnesty |

Examples of Potential Outcomes within each Peacebuilding Theory of Change

| Type of Change | Reduce Resources for War | Work with Elites | Mobilize Grassroots | Healthy Relationships | Address Root Causes | Reduce Violence | Individual Change | Institutional Change |
|---------------------------------------|--|--|---|--|---|--|---|---|
| <i>Change in Attitude</i> | Status is accorded to people with wisdom rather than weapons | Increase elite's appreciation of accountability | Authorities increasingly validate grievances | Reduction in fear of the other | Recognition that the past and the present are different | Belief that respect is derived from appreciation rather than fear | Recognition of own responsibility in the matter | Security forces perceive civilians as people needing their protection |
| <i>Change in Skills</i> | Registration and tracking of small arms and light weapons improved | Improved communications with constituents | Improved advocacy skills | Improve management of rumors | Increased use of non-judgmental language | Able to satisfy own interests through negotiation | Increase in depth of analysis of the conflict | Collaborative, community policing |
| <i>No Change; Maintain Status Quo</i> | Limits on arms manufacturers will be preserved | Decentralization will not be rescinded | Lead activists continue to be able to travel freely | Continue to respect cultural differences | Continue to keep the issues alive in the media | Contain violence to current areas | Persist in advancing own interests | Continue to engage civil society |
| <i>Prevention</i> | Confiscated weapons will be destroyed and not re-enter the market | Will not make decisions in isolation or without consultation | Intimidation by authorities increasingly less effective | Will avoid using stereotypes | Prevent a new cycle of violence from erupting | Prevent additional incidents from escalating into violence | Ensure there are no repercussions | Bureaucrats will not misuse resources |
| <i>Change in Relationship</i> | Collusion in illicit behaviors to legal and transparent relationship | Elites more accessible to those they claim to represent | Increase in intra-group unity | From suspicion to trust | Change in relations between principle, such as opposition party leaders | Police/community relations shift from control-based to service-based | Individual relationships based on dignity and respect | Introduction of citizen consultative processes in key ministries |

There is very little research to guide us in terms of knowing which theories of change and which types of change are likely to be the most effective in identity conflicts, for example. To complicate matters, the issue is not so much about determining which theory or type is the best, but rather when to use each of the many options. Quality monitoring and evaluation can contribute significantly to our understanding of what works and how change happens.

Peacebuilding is about change. As peace builders, we need to be observant enough to see change when it happens, aware enough to understand how change happens, innovative enough to create change, and strategic enough to create change where it can make a difference. Monitoring and evaluation have much to contribute in strengthening our capacity to work with change.

Advanced Concept

Developing Adaptive Change Processes

Programs that focus on one theory or type of change are important; however, they are rarely sufficient to independently foster sustainable societal change. Ultimately, peacebuilding aims to develop change processes that can evolve to meet the challenges of new conflicts, ensuring the continuation of peace writ large. In this manual, we refer to it as the adaptability of change (see Evaluation Objectives) meaning the ongoing dynamic ability to meet new needs, interests, and conflicts in a changing environment. John Paul Lederach frames the challenge as follows:

I should like to posit for the reader that the real world is one of constantly shifting environments and constant adaptation to these shifts. This is particularly true in the setting of deep-rooted conflict and violence. The most realistic, as in the most realpolitik, thing we could do in peacebuilding would be to create processes with peripheral vision, capable of maintaining purpose while constantly adapting to the difficult and shifting sands and tides they must face to survive. The least realistic thing we could do would be to devise rigid processes of politics and social change that are incapable of adaptation. (Lederach, 2005)

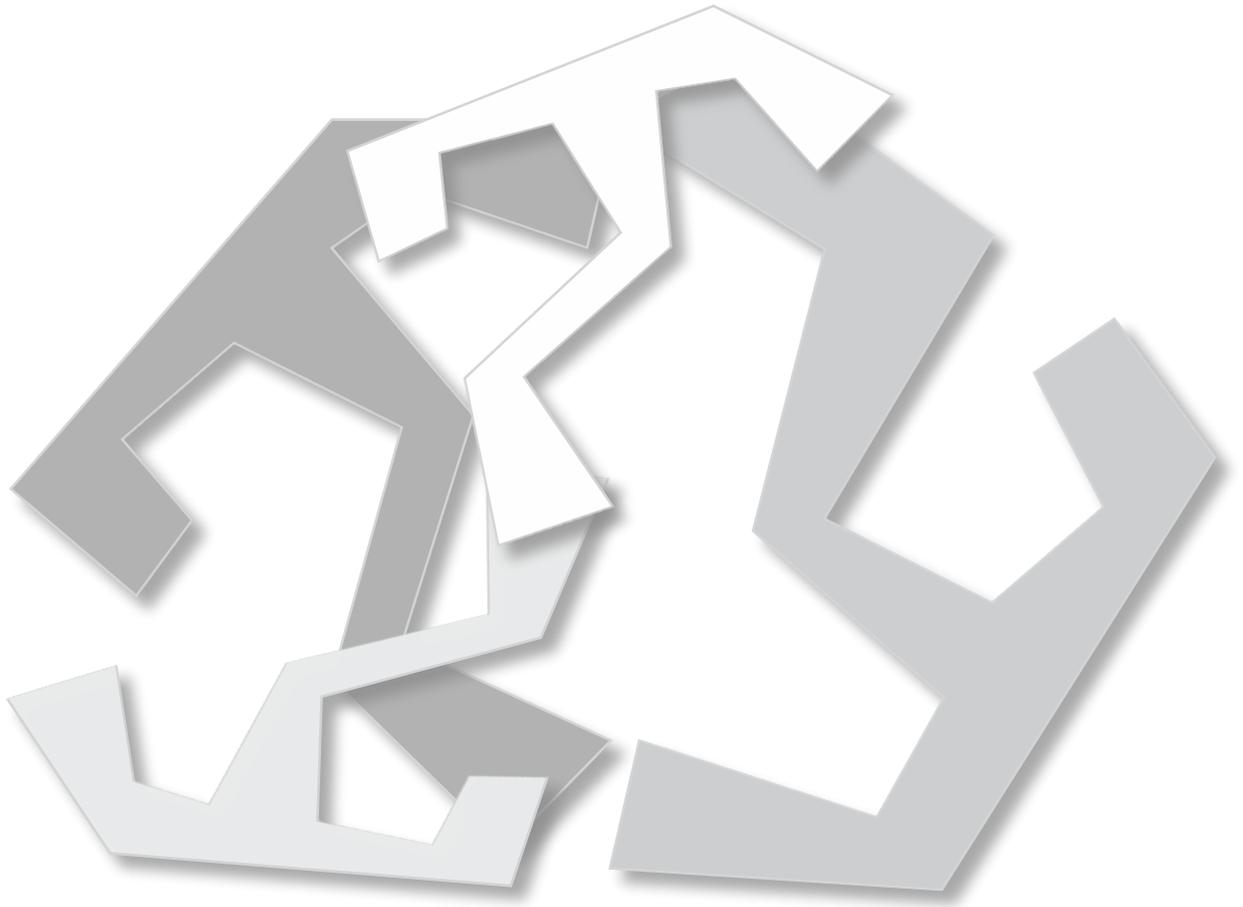
*These adaptive change processes offer the greatest hope of not having to create and recreate new programs with every new conflict. For more information on working with change processes, read Lederach's **The Moral Imagination**.*

Further Reading

John Paul Lederach, *Building Peace, Sustainable Reconciliation in Divided Societies*, USIP, 1997.

John Paul Lederach, *The Moral Imagination: The Art and Soul of Building Peace*, Oxford University Press, 2005.

Malcolm Gladwell, *The Tipping Point: How Little Things Can Make a Big Difference*, Little Brown and Company, 2000.



DESIGN **PROGRAM**

This chapter contains:

1. Techniques promoting creativity
2. Goal setting
3. The design framework
4. Dealing with “donor-speak”

First Steps in the Logical and Results Frameworks

INTRODUCTION

“Some people reach the top of the ladder only to find it is leaning against the wrong wall.”

– ANONYMOUS

Program or project designs that are not closely linked to a thorough and updated conflict assessment risk becoming wonderful ladders leaning against the wrong walls. Unfortunately, in the peacebuilding field, few ladders can be easily moved to another wall. We need to build each ladder for the wall we hope to climb. The first step in program design is to conduct or (re)read the most up-to-date conflict assessments and analyses.

Program or project designs that are not closely linked to a thorough and updated conflict assessment risk becoming wonderful ladders leaning against the wrong walls.

This chapter introduces the basic components of program design, which are the basis for the main design frameworks including the logical and results frameworks. It focuses largely on the design hierarchy, or the relationship between the different levels, with peacebuilding in mind. It also contains a table that shows a comparative illustration of the design language used by a number of the main peacebuilding donors.

This chapter also looks at integrated peacebuilding programming, with lessons from the Collaborative for Development Action. Reference to the design guidelines of specific donors is listed at the end of this chapter.

Why is design important?

The effective use of program design frameworks helps to:

- Improve program effectiveness
- Promote program continuity over time
- Facilitate modification and adjustment to a changing context, and
- Facilitate useful monitoring and evaluation.

How do we come up with creative peacebuilding program designs?

Coming up with the ideas that give life to a peacebuilding initiative is no small feat. Design is inherently a creative process, and in peacebuilding it often requires collaboration. We have yet to develop a guaranteed process, but have found the following practices to be helpful.

TALK, LISTEN, AND ASK DIFFICULT QUESTIONS OF THE STAKEHOLDERS, PARTICULARLY WOMEN AND YOUTH.

Robert Chambers created a very successful development career around one simple concept: the reality that matters most is that of the stakeholder(s). In addition to being best informed about their own reality, they often have invested the most time in thinking about how to change that reality. Insiders' ideas and outsiders' resources can produce surprising results.

CONVENE PEOPLE TO ENVISION A COLLABORATIVE FUTURE.

At one point in the Burundi peace process, many civil society peace workers had become stymied. There could be no further work, they reasoned, until a ceasefire had been negotiated. To them the process had gotten out of sequence. In a workshop anticipating what could be done once a ceasefire was in place, many participants refused to entertain the notion of peace – it wasn't yet a reality. The emphasis of the gathering, many thought, should be on obtaining a ceasefire. These skeptics finally relented and developed a list of potential next steps following a ceasefire. The facilitators then asked why these could not be implemented in the absence of a ceasefire. There were a few steps that would have to wait, but most were things people could start working on immediately. People left the workshop reinvigorated.

CONSIDER THE COUNTER-INTUITIVE. Doing the unexpected can produce positive results. This might mean accepting a proposal at face value or showing up where one is not expected, à la Anwar Sadat, the late Egyptian leader, who traveled unexpectedly to Israel in 1977. It might mean going in the direction of the flow of energy rather than offering resistance, much like Aikido, a martial art based on the use of principles of nonresistance as a way to undermine the strength of the opponent. Agreeing that someone's perspective has merit can take a lot of venom out of an argument.

MAKE BRAINSTORMING A DISCIPLINE. Convene brainstorming sessions with people skilled in lateral thinking and defer judging or assessing ideas until you have a rich pool of possibilities. Use visual stimuli or metaphors to develop unusual associations and new ways of looking at things.

PRACTICE OPPORTUNISM. Every event, even the mundane, is pregnant with opportunity for peacebuilding. Recognizing those opportunities comes with practice since not every opportunity requires action. Once opportunities are recognized, they need to be carefully assessed. Opportunities often have longer shelf lives than we expect. Two types of opportunity are particularly powerful: leverage points and synergies. Leverage points are opportunities to achieve either a scale or a significance well beyond the effort required to implement them. Boycotts are an example of leveraging. Many people, each withholding an individual transaction, fuel boycotts as a way to influence suppliers or governments. Synergies are mutually reinforcing dynamics that exceed the sum of the parts. The synergy or linking of peacebuilding efforts between herders and farmers, on the one hand, with development efforts, on the other, to increase the number of, and access to, boreholes can transform decades of hostility and resentment.

RELEASE THE ARTIST IN PEOPLE. John Paul Lederach maintains that “building adaptive and responsive [change] processes requires a creative act; which at its core is more art than technique.... [As peacebuilders] we need to envision ourselves as artists... and [regain] a sense of the art, the creative act that underpins the birth and growth of personal and social change.” (Lederach, 2005)

Once you have an idea and have discussed it thoroughly with people who know and understand the circumstances of your program, your creative concept will need to be transformed into a program design that can be read, supported, and perhaps even implemented by other people. Most of the rest of this chapter has to do with making that idea understandable and credible, and with illustrating its value and worth.

How are designs built?

Imagine that your analysis has identified a need for something located on a very high shelf, seemingly far out of reach. Below the shelf is a series of small platforms, which are also out of reach as seen on page 29. If we could find a way to get onto the platforms, we might be able to reach the item we need. We look around to see what is available to enable us to climb up to a platform. We could build scaffolding or a ladder. We could try ropes or hire a helicopter to lower us onto the platforms. We choose to build a ladder both because we want to be able to go up and down repeatedly and because we can get the materials quickly and inexpensively. We then get busy ordering supplies, taking measurements, learning ladder safety, cutting wood, etc.

Program design requires our thinking to go through a similar series of challenges in getting from one place to another, or from one result

The important thing to keep in mind is how the different levels – activity, output, objective, and goal – interact with each other.

to another, in order to achieve our goal. In this case, the activities involve building a ladder. The output from those activities is a ladder. Our output – the ladder – will provide access to the platforms. The objective is to stand on the platforms. By standing on two or more platforms, we should be able to accomplish our goal of reaching the item on the shelf.

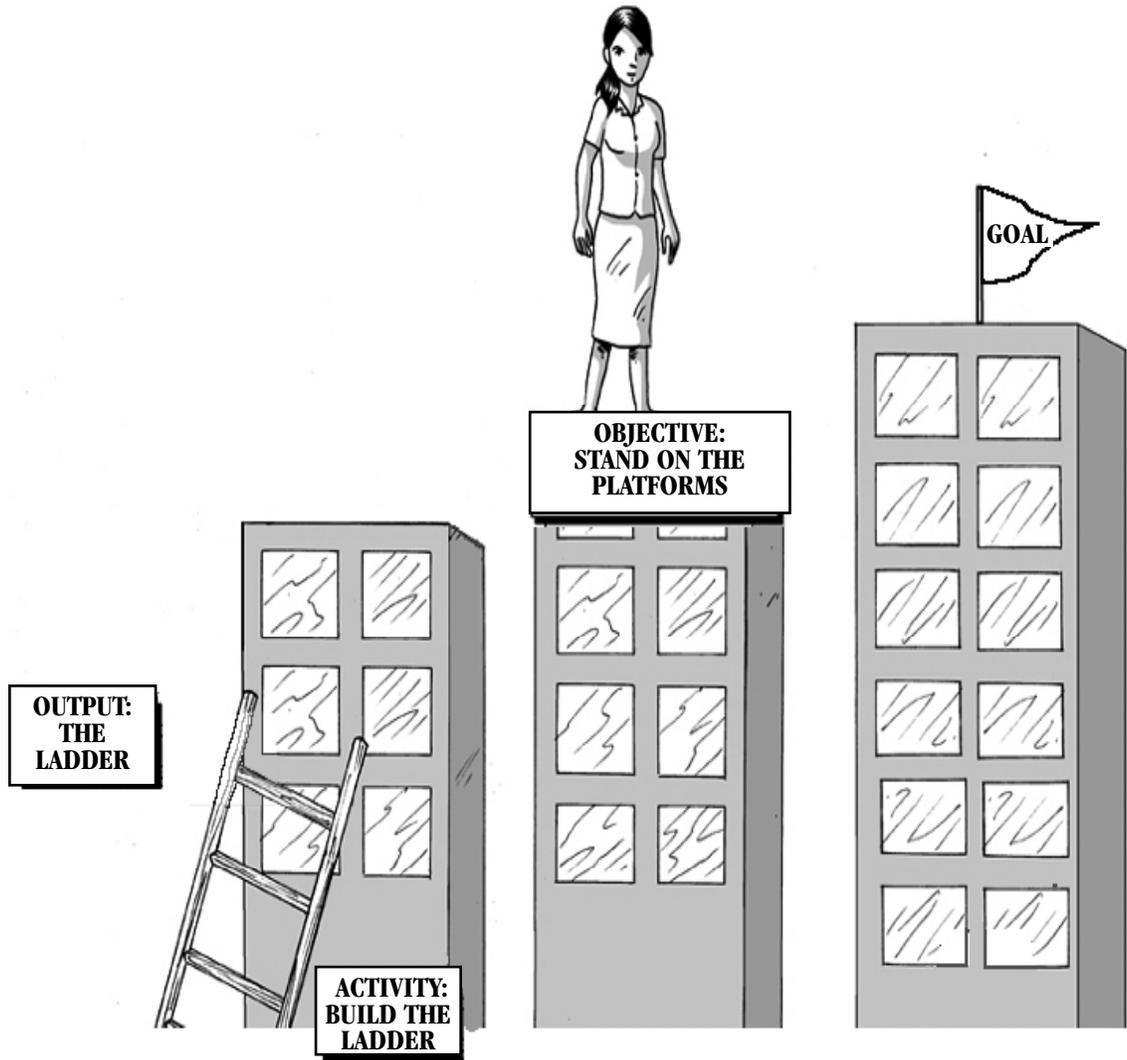


Illustration Representing the Design Hierarchy

The important thing to keep in mind is how the different levels – activity, output, objective, and goal – interact with each other. From the top, looking down, the level immediately below explains how the level above it will be achieved. We will reach the goal by achieving the objectives. We will reach the objectives through the outputs. Finally, we obtain the outputs by implementing the activities.

From the bottom looking up, each higher level explains why we are doing what preceded it on the lower level. We are implementing the

activities to produce the outputs. We need the outputs in order to achieve the objectives. We want to achieve the objectives because they contribute to the realization of the goal.

Ways to Understand the Design Hierarchy

| Why? | Design Hierarchy (levels) | How? |
|---|---|--|
| | Goal: Broadest change in the conflict | How will we achieve the goal? By achieving the objectives below. |
| Why are we trying to achieve these objectives? To achieve the goal above. | Objectives: Types of changes that are prerequisites | How will we achieve the objectives? By producing the outputs below. |
| Why do we need these outputs? To achieve the objectives above. | Outputs: Deliverables or products, often tangible from the activities | How will we produce the outputs we need? By implementing the activities below. |
| Why are we doing these activities? To achieve the outputs above. | Activities: Concrete events or services performed | |

How do we set the goal?

The goal is the broadest change in the conflict that the program hopes to achieve. The goal should derive from the conflict or the people in conflict, rather than from the service or intervention that might be offered. The assessment should help in setting the goal. If the conflict assessment indicates that fear is the major factor contributing to the violence, the goal should focus on overcoming fear or increasing trust. If the assessment reveals that there is an active and high incidence of daily violence against a vulnerable population, a protection goal might be most strategic. Assessments and theories of change are the most common inputs into developing goals.

One way to determine whether or not one has reached the goal is by asking “why” five times. Strengthening the capacity of the office of the National Mediator, for example, is not a goal because it does not describe an intended change in the conflict. Why strengthen the capacity of the National Mediator? The answer to the question helps reveal the goal. In this case, the thinking is that the National Mediator can play an important role in resolving inter-community disputes over land tenure. Therefore, a more accurate goal is to

The goal is the broadest change in the conflict that the program hopes to achieve.

One way to determine whether or not one has reached the goal is by asking “why” five times.

“resolve inter-community disputes over land tenure.” Why do we want to resolve inter-community disputes over land tenure? We want to reduce the incidence of inter-community violence. Why do we want to reduce inter-community violence? We want to achieve greater human security. Why do we want to achieve greater human security? And so forth. Asking and answering the question why will take you to the goal.

Setting the goal at the appropriate level is critical. The team should attempt to identify a goal that is ambitious enough to represent an important change and realistic enough to be achievable given current realities and constraints. In the example above, the goal might be to reduce inter-community violence stemming from land tenure disputes.

The shift toward capacity building in recent years has been driven, in part, by the emphasis on sustainability. Obviously, it would be better to build local capacity to resolve future disputes that evolve in a changing environment than it would be to resolve one isolated dispute. By the same token, it would be better to sustain the capacity to prevent violent conflict rather than to intervene to prevent a single incident. In capacity building programs, the means (i.e., improved services) frequently become the goals, with little concern for the larger conflict that motivated the push for expanded services in the first place. In addition, local capacities often tend to focus on specific types of disputes such as herder/farmer conflicts or community-based disputes rather than those dealing with structural inequities, impunity, identity-based conflicts, and more “intractable” conflicts. The liability of the capacity-building approach is that it is all too common to find that the only change that is tracked relates to the capacity rather than the conflict.

How do we manage goals set by the donor?

Some donors have preferences. For example, the Office of Transitional Initiatives of USAID wants to set goals that are achievable by the implementing partner. Thus, make sure that you know your donor and modify your proposal accordingly. Other donors prefer making peace the goal. Reflecting on Peace Practice, or RPP⁶, calls this goal “*peace writ large*.” Here the idea is that all peacebuilding programs should be able to demonstrate their contribution to peace at the most macro level even if this is not the stated goal of the program. This does not imply that every program needs to be able to achieve peace *writ large*, rather, it indicates that each program can work on its piece of the peace without having to do everything. Partners and other organizations each make specific and coordinated contributions toward peace *writ large*.

Some donors have their own design hierarchies for a conflict and expect their implementing partners, such as NGOs, to make proposals

⁶ More information on RPP can be found online at <http://cdainc.com>.

that dovetail with their design framework. This requirement often means that the lowest level of the donor's framework becomes the highest level of the local implementing partner's program design structure.

Donors also need help in setting goals, however. Don't casually discard your analysis and planning to blindly integrate the donor's goal. In many cases, goals and objectives can be negotiated to meet mutual interests or simply be submitted through another channel, such as an unsolicited proposal.

Project objectives describe the types of changes that are prerequisites to achieving the goal.



“Dear Mr. Gandhi, We regret we cannot fund your proposal because the link between spinning cloth and the fall of the British Empire was not clear to us.”

Written by *M. M. Rogers* and illustrated by *Ariv R. Faizal, Wabyu S., Ary W.S.*
Creative team for Search for Common Ground in Indonesia

How do we define the objectives?

Project objectives describe the types of changes that are prerequisites to achieving the goal (see the chapter on Understanding Change page 18 for examples of types of change). If our goal is to increase the level of trust between two communities, one objective might be to increase the amount of constructive, safe contact between those communities. If we are seeking to protect vulnerable populations, an objective might be to expand and improve policing efforts.

Strengthening the offices of the National Mediator is an objective because it does not involve a change in the conflict. It is one means of achieving the goal of reducing inter-community violence stemming from land tenure disputes. There are generally several (1-3) objectives to be achieved under any given goal. Other possible objectives in the National Mediator example might be to introduce a land title and registration system, adjudicate pending cases from earlier failed land redistribution initiatives, or broadcast harvest and migration schedules.

Objectives ensure a logical link between activities and goals, which is why they can be difficult to develop.

Objectives ensure a logical link between activities and goals, which is why they can be difficult to develop. As peacebuilders, we want peace. As practitioners, we have creative and exciting ideas about activities and events. Determining and demonstrating the connection between peace and our proposed activities is critical.

How do we select the activities and outputs?

Activities are the concrete events or services that program staff members and participants implement such as dialogue sessions, mediations, exchange visits, curriculum development, radio soap opera production, community organizing, sporting events, training, negotiations, etc. The immediate deliverables or products from activities, which are often tangible, are called outputs. A training activity will produce trained people as an output. A mediation will produce facilitated negotiation sessions as an output. Recording a radio soap opera will produce radio shows.

Those well-trained in implementing activities correctly point out that activities also have goals, objectives, and activities within the implementation of those activities.

Outputs and activities are so closely linked that they are best conceived together rather than separately. Well-produced outputs contribute to the achievement of the objectives. Outputs are like the rungs on the ladder – each one should move us closer to the corresponding objective.

Activities are the concrete events or services that program staff members and participants implement.

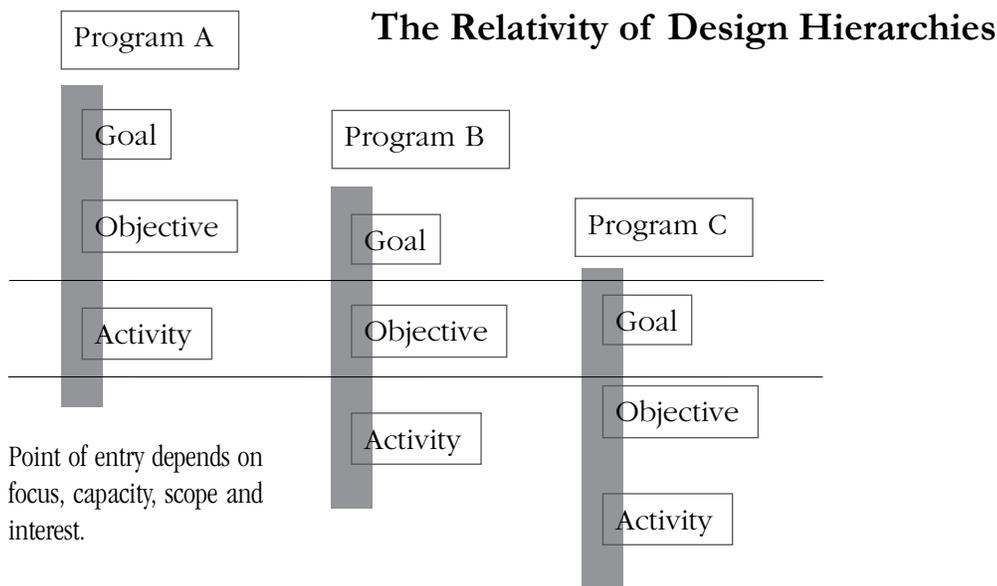
The challenge in selecting activities is making a few strategic and effective choices from a large menu of possibilities. Criteria established prior to assessing the many possible activities will help weed out the great ideas of little importance. Consider the outputs that each activity will generate. Are they needed? Are they enough to lead to the change anticipated in the objective? Making a contribution toward the objectives has to be among the top criteria in choosing among multiple activities.

Aren't all design hierarchies relative?

Those well-trained in implementing activities correctly point out that activities also have goals, objectives, and activities within the implementation of those activities. For instance, an individual training, which is an activity within a larger program, contains a goal for that training, objectives on how to reach that goal, and a set of activities that collectively

constitute the entire training. This “relativity,” to borrow from Einstein, is extensive. Three separate programs, each with its own design, could be focused on the same intervention at a different level in their design. What might be characterized by a donor – such as DFID, for example – as an activity might be a goal for a local peacebuilding organization.

Understanding relativity in program design can be instrumental in identifying partners, building integrated programs, finding points of collaboration, reinforcing interventions, and creating synergies to leverage greater results.



Assumptions are the unproven connections between levels.

How do we identify assumptions?

Assumptions are the unproven connections between levels. Most assumptions focus on:

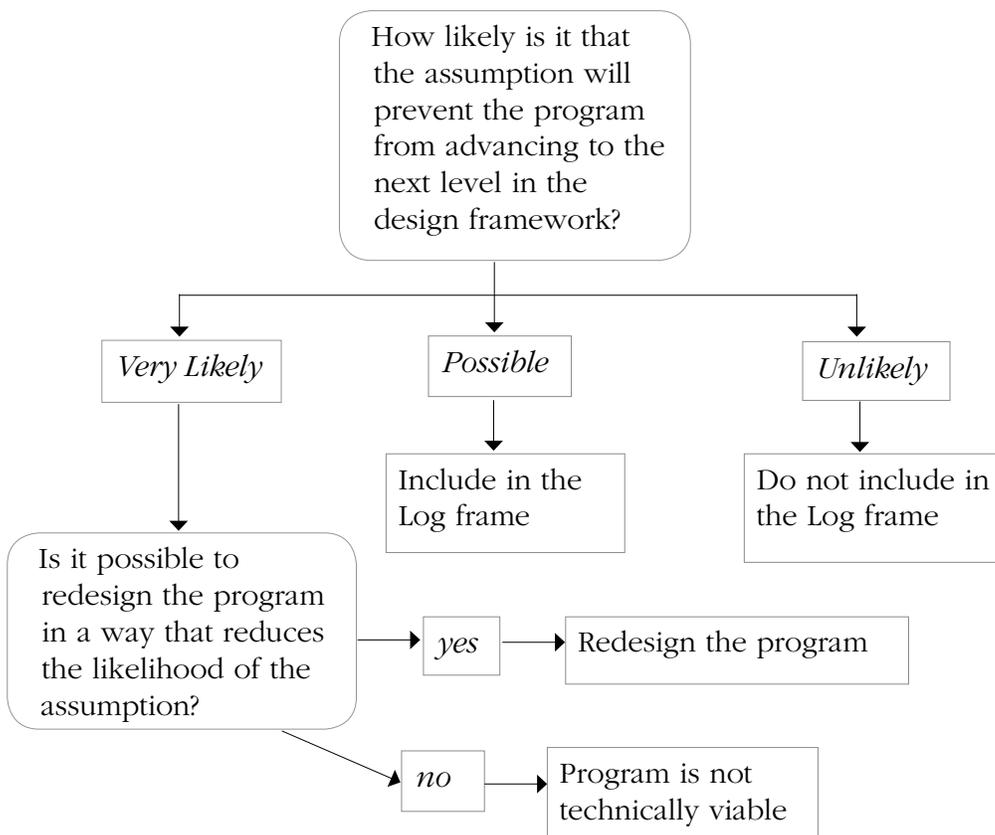
- How the context will evolve.
- Program philosophy or approach.
- Participation. In a war zone, for example, many programs assume there will be sufficient security to safely access the people or certain areas.
- Our understanding of how things work in life. For example, we assume that relationships built by enemies in the safety of a workshop or a structured exchange will enable them to behave differently upon their return to their everyday, often-polarized environments. We assume that greater transparency will lead to better governance, rather than to well-publicized corruption. We assume that a ceasefire opens space for negotiations rather than serving as a time for restocking ammunition and reinforcing military positions.

It is important to identify assumptions early in the design process, ideally when establishing the goals and objectives, and certainly before developing indicators and the means of verification.

It is important to identify assumptions early in the design process, ideally when establishing the goals and objectives, and certainly before developing indicators and the means of verification. If the assumptions are unlikely to hold (remain true), the goals and objectives may need to be further refined. Once the effort has been made to develop indicators, people are reluctant to alter them because of an inaccurate assumption. Therefore, develop assumptions for each level and determine if the design hierarchy is still viable before proceeding to indicators.

When we begin to look at assumptions we quickly become swamped. Which assumptions matter the most? The following flowchart can help in determining which assumptions to include in the design. Generally speaking, we need to consider those assumptions that could possibly prevent or block a connection between one level in the design and another level.

Determining Assumptions



Consider a peace media program using radio in which there is an assumption that accurate and balanced information will encourage people to deal with conflict without resorting to violence. In an area of low radio listenership, sporadic electricity, and high population mobility, this assumption could very likely block the program from achieving its objective of reduced violence. Broadcasts are frequently interrupted by power cuts and few people listen to the radio because they are often

moving around to ensure their safety. Consequently, the program should be redesigned in such a way that additional activities are directed at changing the behaviors of key stakeholders, such as conducting mobile training workshops for leaders, organizing zones of peace, and distributing solar-powered radios.

This assumption might be different in an area where listenership is high and where past evaluations reflect evidence of significant behavioral change among certain segments of the population which have repeatedly been exposed to balanced and accurate media. We would, therefore, include this assumption in the design. The same program assumes that people will be able to purchase batteries to listen to the radio. In this specific context, such an assumption is very unlikely to prevent the program from advancing because batteries are cheap, produced locally, and in large supply. Thus, we would not include this assumption in the design.

It is important to identify and determine how to deal with assumptions before beginning to work on indicators and other parts of the monitoring and evaluation plans.

A single goal promotes clarity. There is no fixed, required number of objectives, although 2-3 objectives are the norm.

How do we put all of these ideas together?

A single goal promotes clarity. There is no fixed, required number of objectives, although 2-3 objectives are the norm. There is also no set number of activities for each objective. The activities and outputs, when viewed together, should reflect a convincing mass or momentum that will result in the planned objective.

Putting It All Together

| | Narrative Summary | Assumptions |
|--------------------|---|--|
| Goal | Reduce the incidence of youth violence in five counties by 40% in three years | Youth violence is largely due to alienation, isolation and weak social connections with the larger community |
| Objective 1 | Increase the safe, no-violence areas | Perpetrators will continue to respect no-violence zones |
| Objective 2 | Promote collaborative relationships between major youth groups | Bridges between local youth groups will reduce the amount of inter-group violence |
| Objective 3 | Improve relationships between communities and youth groups | Improved relations between youth and adults will contribute to a reduction in violence |

Putting It All Together

| | Narrative Summary | Assumptions |
|--------------------|--|---|
| Activity 1a | Organize an additional 35 safe sites in addition to the existing 12 no-violence zones | Police can expand coverage to fully patrol additional sites |
| Activity 1b | Introduce community policing in heavily impacted neighborhoods | Assumes neighborhood will participate in community policing |
| Activity 2a | Train 120 young leaders in non-violent communication and conflict resolution skills | Assumes youth leaders will apply their training to local conflicts |
| Activity 2b | Provide financial and programmatic support to youth groups | Assumes additional support will help youth groups attract new members and retain existing members |
| Activity 2c | Organize youth forums about non-violence and the creation of no-violence zones | Assumes forums can be constructive means for youth to express needs and interests |
| Activity 3b | Establish advisory commissions of young people attached to local government offices that work on issues of interest to youth | Local politicians and authorities will consider proposals from youth in exercising their duties |

The concepts outlined above – hierarchies and assumptions – are common to both logical and results frameworks

The example above is an excerpt from the logical framework, minus the monitoring and evaluation components to be discussed later. When completing a logical frame, complete these two columns first before proceeding to the monitoring and evaluation columns.

How are logical and results frameworks different?

The logical framework is oriented toward work to be done in the future. The results framework describes the results as if the program has already been completed.

The concepts outlined above – hierarchies and assumptions – are common to both logical and results frameworks, which are the most common design frameworks in the peacebuilding field at present. These frameworks are intended to facilitate design and enable better management of project implementation. As such, they are actually management tools, but for the purpose of this manual, we will refer to them as program or project design frameworks.

The major difference between the two versions – logical and results – is one of perspective. The logical framework is oriented toward work to be done in the future. The results framework describes the results as if the program has already been completed. Because the results

framework focuses on achievements, the logical connections between levels are sometimes clearer and easier to understand. The table below illustrates how each framework handles the same objective.

Results vs. Logical Frameworks

| Logical Framework | Results Framework |
|---|---|
| Objective | Intermediate Result |
| To reduce by 50% within three years the number of people in each identity group who fear the other identify group | After three years, 50% of the people in each identity group no longer fear people of the other identity group |
| Both ethnic groups will respect traditional cultural heritage events of the other group | Cultural heritage events of both groups were celebrated in an atmosphere of acceptance and respect |

For additional information on how to complete specific design frameworks, see the resource guide at the end of this chapter.

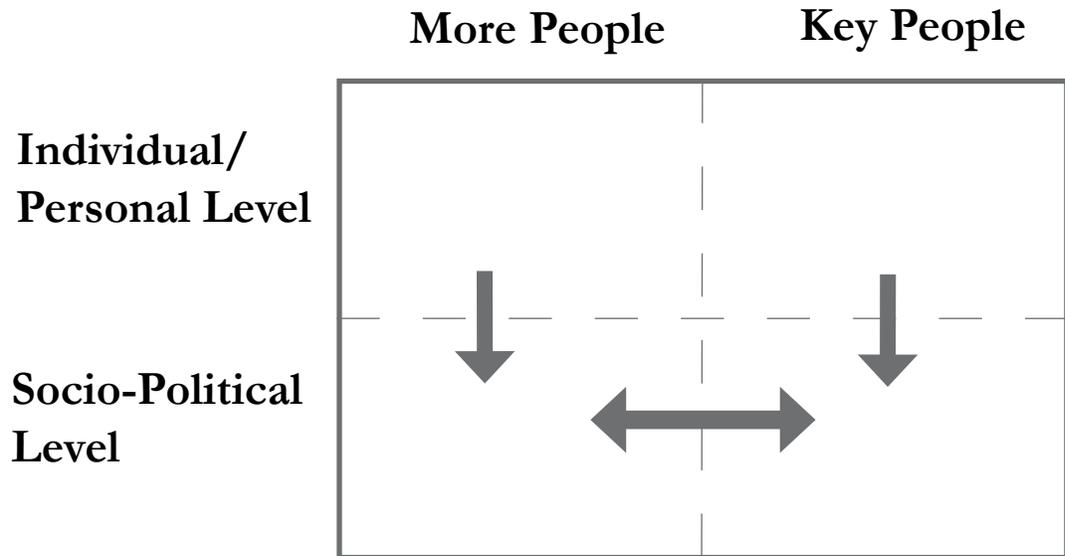
Advanced Concept

Designing Integrated Peacebuilding Programs

Integrated programming is often perceived as combining programs from different sectors, such as health and peacebuilding or education and peacebuilding. For our purposes, integrated peacebuilding refers to the connection between two or more approaches to peacebuilding in an effort to promote synergies and leverage greater results. This is also referred to as vertical integration. Another way to think about this concept is in terms of integrating multiple theories of change. More often it is the realization that changes require multiple stakeholders, with differing interests, to agree on a variety of solutions or changes. Accessing each stakeholder may require a different approach to each. Programs working with all the needed stakeholders or those linked to other programs that include all needed stakeholders are likely to succeed.

In Confronting War, Critical Lessons for Peace Practitioners, the authors have synthesized 27 case studies into several concrete strategies for integrated peacebuilding programs. They first categorized all the programs into four types along two axes. One axis describes the targeting of people, ranging from more people to key people. The other axis looks at the locus of the change ranging from individuals to institutions.

Interconnections among approaches and levels



Arrows indicate connections that leverage results

In sum, the findings suggest programs will be more effective when they:

- **CONNECT THE INDIVIDUAL/PERSONAL LEVEL AND THE SOCIO/POLITICAL LEVEL**. Evidence shows that when programs focus only on change at the individual/personal level without regard to how these may be translated to the socio/political level, actions inevitably fall short of having an impact on the larger goals. Many peace efforts that work either with more people or with key people at the individual/personal level aim to build relationships and trust across lines of division, to increase tolerance, to make peace seem possible and within reach to people, or to inspire hope. Practitioners and communities talk of having been “transformed personally” by a particular program or “having my perceptions about the other side changed” or “improving my relationships and communication with individuals on the other side.” The evidence shows, however, that in order to have a real impact on conflict, personal change must be translated into actions at the socio/political level.
- **CONNECT MORE PEOPLE AND KEY PEOPLE AT THE SOCIO/POLITICAL LEVEL**. Evidence shows that even in activities at the socio/political level, work with more people is not enough if it does not reach key people, while work with key people is not enough if it does not reach more people. Some examples will illustrate these

common problems of peace programming. An agency organized an ongoing high-level dialogue process involving influential people with decision-making power in the official negotiations. This resulted in improved communication in the official negotiations and the uptake of some ideas on solutions. However, after several years, the two sides remained far apart on a political resolution. Leaders on both sides claimed they were blocked from making more progress. Public opinion was described as “not ready.” The effort was stuck at the key people level, and was unable to affect the more people level.⁷

How do we monitor and evaluate if we don't have a design?

The original design document and its subsequent incarnations are important references in monitoring and evaluation. Well-written plans, such as logical and results frameworks, that contain useful indicators can play a significant role in facilitating implementation, monitoring, and evaluation, contributing to learning, and cutting costs. In the absence of an initial design, evaluators may need to work with practitioners to recreate the project's/program's original goals and objectives. When this type of effort is undertaken, the result is more likely to reflect current perspectives rather than those held at the outset of the program. Placing emphasis on the goals and objectives sought today can be problematic because there is a risk of overlooking important developments and modifications made either after the original design or during implementation. Subsequent developments may contain important strategic breakthroughs, missed opportunities, or flawed decisions made on false assumptions.

Most donor agencies and actors within the peacebuilding world use a core set of design terms; however, they use them in vastly different ways.

If the donors use design terms differently, how do we know what they mean?

Most donor agencies and actors within the peacebuilding world use a core set of design terms; however, they use them in vastly different ways. In other words, “impact” to one donor may mean “results” to another, while “goal” for one organization may refer to “overarching objective” to another. These hidden differences cause confusion and frustration for those seeking to translate their project design into the appropriate format for each donor.

The OECD (Organization for Economic Cooperation and Development) has produced a glossary of terms for evaluation that is becoming widely adopted by evaluation departments within donor governments (available online at <http://www.oecd.org/dataoecd/29/21/2754804.pdf>). However,

⁷ Mary Anderson and Lara Olsen, with assistance from Kristin Doughty, *Confronting War: Critical Lessons for Peace Practitioners*, Collaborative for Development Action, 2003

these definitions have not yet been transferred into the design frameworks for most donor agencies, which means that the glossary cannot be relied upon for donor design purposes.

To assist the need for translation, a terminology decoder for donor design, monitoring, and evaluation terms has been developed, found on page 42. Since donor frameworks and requirements change frequently, check with the donor to ensure that this language is still current before submitting a proposal for funding.

Please see Appendix A on page 227 for a list of the source documents for each of the donor agencies.

Further Reading

USAID Center for Development Information and Evaluation, *Performance Monitoring and Evaluation Tips: Building a Results Framework*, Number 13, USAID, 2000.

Innovation Network, Inc., *Logic Model Workbook*, 2005.

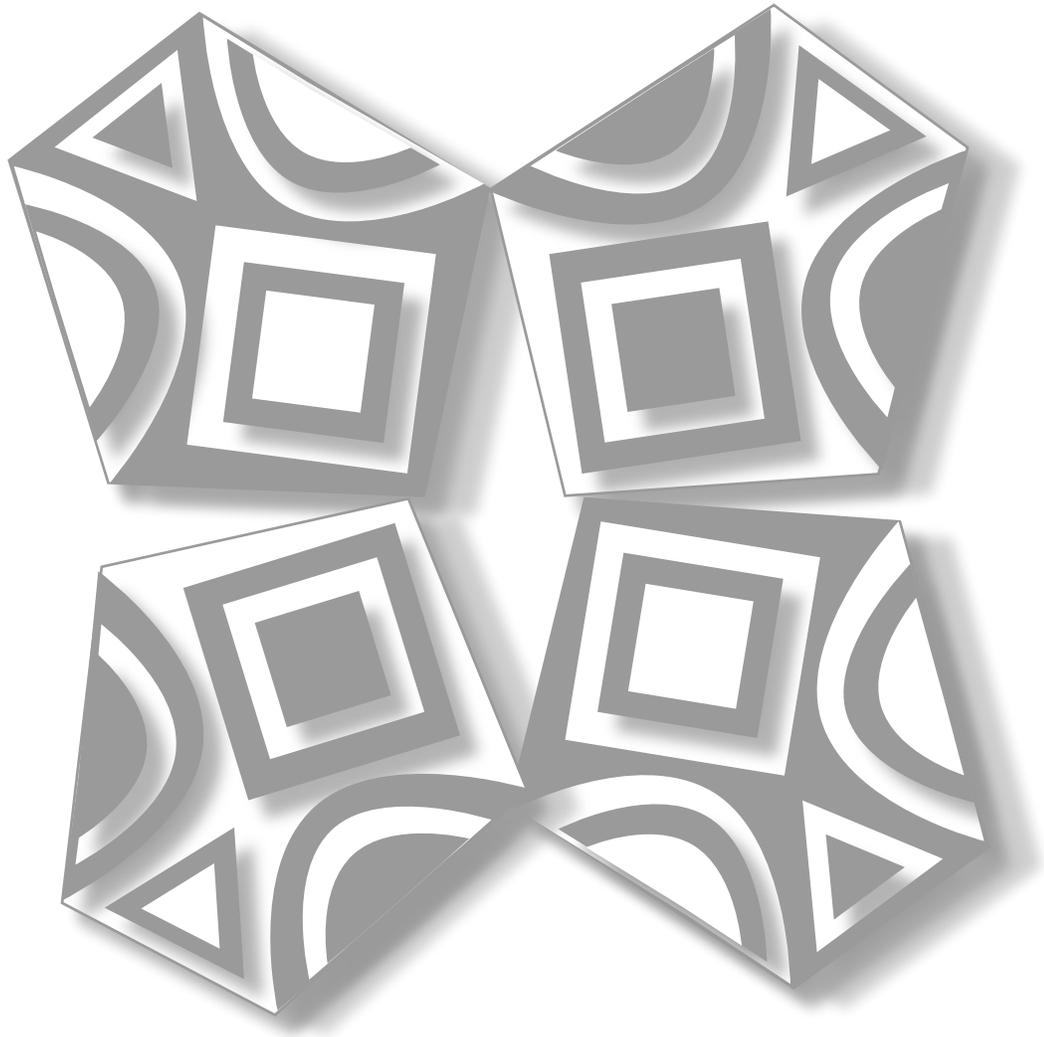
http://www.innonet.org/index.php?section_id=62&content_id=143

Organization for Economic Cooperation and Development, *Glossary of Key Terms in Evaluation and Results Based Management*, 2002.

<http://www.oecd.org/dataoecd/29/21/2754804.pdf>

Donor Terminology Decoder

| SFCG | Project Narrative | | | | Results | | |
|---------------------|---|---|----------------------|---------------------|---------------------|---------|--------|
| | Goal | Objective | Activity | Inputs | Output | Outcome | Impact |
| UNDP | Outcomes | Outputs | Activities | Inputs | --- | --- | -- |
| CIDA | Project Goal (Program Objective) | Project Purpose | Resource | Resource | Output | Outcome | Impact |
| EuropeAid | Overall Objective | Purpose | Activities | -- | Results | -- | -- |
| TACIS/ EuropeAid | Overall Objectives | Specific Objectives | Activities | Means | Expected Results | | |
| DFID | Goal | Purpose | Activities | | Output | | |
| World Bank | Goal | Development Objective | Component Activities | Input/ Resources | Output | Impact | -- |
| UNHCR | Goal | Objective | -- | Input | Outputs | Impact | -- |
| SIDA | Goal/Overall Objectives/ Development Objective | Project Purpose/ Immediate Objective | Activities | Resources/ Input | Results/ Outputs | -- | -- |
| RELEX | Overall Objectives | Operation Purpose | Activities | -- | Results | -- | -- |
| SADC | Overall Goal | Project Purpose | Activities | -- | Results or Outputs | -- | -- |
| EIDHR | Overall Objectives | Specific Objectives | Activities | Means | Expected Results | -- | -- |
| Danida | Development Objective | Immediate Objective | Activities | Inputs | Outputs | -- | -- |
| USAID | Strategic Objective | Immediate Results | Activities | | -- | -- | -- |



INDICATORS

This chapter contains:

1. Descriptions of quantitative and qualitative indicators
2. The components of indicators
3. Ways to develop indicators
4. Indicator quality checklist

INTRODUCTION

“The only difference between stumbling blocks and stepping stones is the way in which we use them.”

- ADRIANA DOYLE

This chapter explains the purpose of indicators as a means of measuring change. It focuses on two of the most common types of indicators: quantitative and qualitative indicators. It describes the basic components of an indicator and outlines some very general suggestions for developing new indicators. It also includes important dimensions for analysis that need to be considered in creating indicators and very real risks in becoming indicator-driven.

What is the purpose of an indicator?

Indicators consist of information that signals change. An indicator is a quantitative or qualitative factor or variable that provides a simple and reliable means to reflect the changes connected to an intervention. Indicators enable us to perceive differences, improvements or developments relating to a desired change (objective or result) in a particular context. “Indicators are inevitable approximations. They are not the same as the desired change, but only an indicator of that change. They are imperfect and vary in validity and reliability.”⁸

Where the desired change is concrete, tangible, and measurable, indicators are not needed. If the intended output was 500 brochures, no indicator is needed – simply count the number of brochures produced. Consider the example from a six-month long peace media radio program, in a context where inaccurate rumors often cause violence. The objective of the project is to increase the public’s access to accurate information within 24 hours of when the rumor that promotes violence started circulating. The desired change of substituting rumors with accurate information is concrete, tangible, and directly measurable. There is no need for a separate indicator.

Where the intended change is more abstract, indicators help approximate the change. For example, in order to monitor a change in the level of trust between groups, one might look at child care practices to see if adults from one group are permitted to care for children from the other group. To detect changes in equality one might monitor inheritance, land ownership, and employment.

An indicator is a quantitative or qualitative factor or variable that provides a simple and reliable means to reflect the changes connected to an intervention.

⁸ Michael Patton, *Utilization-Focused Evaluation*, Third Edition, SAGE Publications, 1996, pp.159.

Indicators are used in establishing baselines, monitoring, and evaluation.

In peacebuilding, indicators enable us to work with many intangible issues that are often at the root of the conflict.

Indicators are used in establishing baselines, monitoring, and evaluation. Information is gathered in the baseline to set the target for the indicator. Indicators can then be used for determining progress toward results in monitoring as well as in monitoring the context of the conflict. For example, peace activists often track changes in militarization in order to be able to anticipate changes in the conflict context. If we want to know about changes in militarization, military recruitment is one good indicator. Recruitment may signal a number of different things: the replacement of an aging force, an increase in soldier/officer ratios, or a more equitable regional representation within the military. A more complete picture is obtained by adding additional indicators such as defense spending, force deployment, and arms purchases.

What changes in recruitment tell us depends on who we are and what we want to know. For some people, increasing recruitment may not represent the most important thing to know about militarization. Arms suppliers may want to know about anticipated demands for additional weaponry. Officer training academies may want to know the rate of recruitment in order to schedule officer-training programs accordingly. Bilateral aid agencies may want to ensure that their assistance is not being misused. Neighboring countries want to ensure that their relationships and their security are not in jeopardy.

Whatever recruitment signals to us, it does not tell us why there is an increase in militarization. Recruitment as an indicator cannot explain the complex governance and security issues behind the choice to militarize. At best, this indicator tells us that a change we are interested in is happening. Indicators cannot explain why or how that change occurs.⁹

In peacebuilding, indicators enable us to work with many intangible issues that are often at the root of the conflict. Success in selecting and developing good indicators is directly related to the depth of the conflict analysis, the understanding of the context, and expertise in designing effective interventions.

What are the basic components of an indicator?

Indicators need to contain certain basic information and also be able to pass tests of reliability, feasibility (see page 70 under Means of Verification for more information on feasibility), and utility in decisionmaking. The basic, minimal information contained in an indicator is outlined below. Not all this information is needed for each indicator.

⁹ Adapted from *RBM* [Results-Based Management] in *UNDP: Selecting Indicators, Signposts of Development*, (UNDP, 20002).

1. What is to be measured
– what is going to change
2. Unit of measurement to be used to describe the change
3. Pre-program status/state, also known as the baseline (where possible)
4. Size, magnitude, or dimension of the intended change
5. Quality or standard of the change to be achieved
6. Target population(s)
7. Timeframe

The following table illustrates the different components from two example indicators.

Examples of Indicator Components

| | Example 1 | Example 2 |
|---|---|--|
| <i>Indicator Components</i> | Indicator: Increase the percentage of participants from the southern districts reporting an improvement in their relationship with the other(s) to the point where they now enter each other's homes from 20% in 2005 to 70% by 2008. | Indicator: Increase the number of men and women participating in at least two inter-community activities from 75 men and women/year in 2005 to 450 men and women/year for all ten program communities before the end of 2007 |
| <i>What is to be measured</i> – what is going to change | Participants reporting an improvement in their relationship with the other(s) | The number of men and women participating in at least two inter-community activities |
| <i>The unit of measurement to be used to describe the change</i> | Percentage of participants | Number of women and men |
| <i>Pre-program status/state, also known as the baseline</i> (where possible) | From 20% of the participants in 2005 | From 75 men and women/year in 2005 |
| <i>The size, magnitude or dimension of the intended change</i> | To 70% of the participants in 2008 | To 450 men and women/year before the end of 2007 |
| <i>The quality or standard of the change to be achieved</i> | Improved to the point where they enter each others' homes | At least two inter-community activities |
| <i>Target population(s)</i> | People in the southern district | Men and women from all 10 program communities |
| <i>The timeframe</i> | Between 1 January 2005 and 1 January 2008 | Between 2005 and the end of 2007 |

First and foremost, an indicator should concretely specify what is to be measured. In the following example, the indicator is measuring changes in mobility within areas controlled by the other side. There is an assumption that increasing mobility (a change in behavior) is a signal of increasing trust (a change in attitude or thinking).

Examples of Indicators

| Objective | Indicator |
|---|---|
| Increase trust between the two communities. | 50% of men, women, and children from each side increase their mobility within the areas controlled by the other side by at least one square kilometer per year. |

The unit of measurement in this example is square kilometers. It is important to set a target, i.e., to determine the size, magnitude or other dimension of the intended change. The example above contains two targets: “at least one square kilometer” (geographic) and “50% of men, women and children” (demographic). In this example, the indicator includes more specific information on gender and age that was not included in the objective. It has been included to allow program managers and decision-makers to see if, during the course of the project, any one segment of the population has been excluded or needs additional attention in order to achieve the objective.

Consider the following example where three communities have been unable to resolve disputes over shared natural resources, transportation, and garbage disposal. The conflict assessment revealed that past efforts to work together toward solutions always used positional bargaining, which resulted in threats and intimidation.

Examples of Indicators

| Objective | Indicator |
|---|--|
| Increase inter-community collaboration on public policy issues that address common interests. | Expand from twice/year to six times/year the number of public policy debates or forums where all three communities contribute interest-based solutions on natural resource management disputes by the end of 2009. |

In this example, it is assumed that contributing interest-based solutions (a change in process) reflects an increase in effective inter-community collaboration (a change in relationship). The unit of measurement is the number of instances where all three communities contributed interest-based solutions during policy debates. The size or magnitude of

change involves an increase from two debates/year to six debates/year. The target population is three communities and the timeframe concludes at the end of 2009.

This indicator focuses on process or implementation. It can be helpful for monitoring the collaborative process, but it cannot inform the practitioners of any changes at the outcome level. It is possible that the communities will succeed in offering interest-based solutions without coming to agreement on solutions and without increasing collaboration. Given the baseline practice of positional bargaining, there is a good chance that interest-based negotiation could enhance collaboration.

How do we set the targets of an indicator?

The previous example raises the question, “How much change is enough?” This is akin to the question, “What is success?” While there are no hard and fast answers to these questions, there are some basic guidelines.

- Know the size or magnitude at the beginning (see baselines).
- The amount of change needs to be large enough to be significant.
- The amount of change needs to be small enough to be achievable within the means (i.e., budget, staff, and capacity) of the project.
- Review past records and reports for previous experience.
- Ask yourself, “What does that mean in real terms?” For example, it would not be useful to set an arbitrary target such as “50% increase in the number of adolescent boys and girls who complete a peace education course in the province during their fifth year.” In the first year of the project, an increase of 50% of zero would be meaningless.
- Alert the donor that you will need to adjust the targets following the baseline and as you gain experience.
- Adjust the targets after the baseline.
- Adjust the targets after you have experience.

Once the basic components have been determined, potential indicators need to pass three tests before entering into final consideration.

How do we know the indicator will work?

Once the basic components have been determined, potential indicators need to pass three tests before entering into final consideration:

- Reliability: Consistency of the finding regardless of who makes the measurement.
- Feasibility: Ease in collecting the information.
- Utility in decisionmaking: Critical to informed choices.

Consider a security reform program in which one of the objectives is “to increase the accountability of the armed forces.” The following table has

three different indicators for this same objective. Each indicator scores differently on the three quality tests.

Quality Tests

| Proposed Indicator | Reliability | Feasibility | Utility in decisionmaking |
|---|---|--|---|
| Within 12 months, 80% of all officers can cite the types and ranges of sanctions that correspond to the most serious human rights violations. | Involves some sort of test which is likely to yield the same results no matter who applies the test. | Testing a significant sample of officers is only feasible if there is full support and endorsement of such testing by the chiefs of staff. | This helps us understand what officers know and the degree to which ignorance is a factor. In conjunction with other indicators, it also may give us insights into the degree of influence officers have over soldiers. |
| Increase by 50% the number of sanctions from military tribunals that fall within anticipated norms by end of year two. | Unless the norms are stated, this is less reliable because it requires a judgment – does the sanction fall within the norm? | If access to records from military tribunals is difficult to obtain, this indicator is not feasible. | Is military justice a viable deterrent to human rights abuses by the military? |
| Increase by 30 % each year the number of people living near military bases who believe military leadership seriously investigates instances of alleged abuse and prosecutes it accordingly. | If all variables such as time of day, sample size, and selection methods are the same, the reliability should be within acceptable standards. | With both community and base endorsements and the requisite security, this could be feasible. | Are public perceptions changing proportionately to the changes being implemented by the military? |

The following checklist can help in assessing choices, and the value of those choices, among a variety of proposed indicators.

TARGETED

- Element of change: What is changing?
- Target group: Who is involved in the change?
- Location: Where is the change located?
- Timeframe: When is the change to happen?

MEASURABLE

- Specific unit(s) of measurement to be used: What will be measured, counted, weighed or sized?
- Reference to a baseline/benchmark for comparison: What was the measurement at the starting point?
- Qualities are defined: Words like “effective, appropriate, successful,” are defined clearly.

RELIABLE

- Quality of the information is credible.
- Assumptions are minimal, or at least clearly stated.
- Connection between the indicator and what you are trying to prove is direct.
- Everyone collecting the information will find the same thing.

Quantitative indicators are measures of quantities or amounts.

FEASIBLE

- Means of verification is viable and doable.
- Information can be obtained.

Qualitative indicators are people’s judgments or perceptions about a subject.

UTILITY IN DECISIONMAKING

- The information is linked to key decisions.
- The information has major importance in the decision.

What is the difference between qualitative and quantitative indicators?

Quantitative indicators are measures of quantities or amounts. An example would be a 50% increase in the number of people who enroll their children in ethnically mixed schools by the end of the project. Another example would be, “500 disputes resolved by trained mediators over 18 months.”

Qualitative indicators are people’s judgments or perceptions about a subject. An example would be, “25% increase in the level of confidence people have in their ability to circulate safely in all areas in their community by end of project.” Another example might be, “10% decrease in fear of violence in village D in 6 months.” Most qualitative indicators contain a number or numeric components so you need to look beyond numbers to what is actually being measured. Check to see if the change in question relates to some sort of opinion, belief, or way of thinking. If not, it is most likely a quantitative indicator. If it describes the implementation of an activity or a one-off event, it is almost certainly a quantitative indicator.

Examples of Quantitative and Qualitative Indicators

| Objective | Quantitative Indicator | Qualitative Indicator |
|--|---|---|
| Increase solidarity between 450 former enemies in five municipalities in Chalatenango over three years | Percentage of the former enemies in five municipalities in Chalatenango who have joined mixed-community organizations at the end of year one <i>Measures the size of membership in an organization</i> | Percentage of former enemies who, at the end of year one, routinely identify themselves as members of the larger community rather than belonging to one group or faction <i>Measures change in how they describe themselves; a quality of their identity</i> |
| Enhance capacity of regional and local government institutions and communities to monitor, report, and manage conflict in two years in three southern provinces. | Number of disputes reported at each level during the course of the project <i>Measures the incidence of reporting</i> | Number of instances at each level where authorities believe that monitoring reports lead to a timely intervention and the prevention of escalation over the course of the project <i>Measures the authorities opinion of the contribution of monitoring toward intervention and prevention</i> |

Programs combining both qualitative and quantitative indicators often demonstrate a richer understanding of the dynamics at play.

Programs combining both qualitative and quantitative indicators often demonstrate a richer understanding of the dynamics at play. Consider a program whose objective is to increase the political empowerment of women.

The **QUANTITATIVE** indicator is, “50 women elected parliamentarians in the next election.”

While this is a laudable objective and an adequate indicator, alone it does not provide the full story. Add to it one or both of the following qualitative indicators and we begin to see that there are other dynamics at play, such as the ability to exert influence once in office.

The **QUALITATIVE** indicators are:

“10% increase in women parliamentarians’ belief that their voices are making a difference in decisionmaking.”

“15% decrease in elected women’s perception that they are marginalized in decisionmaking.”

What are the other important dimensions for analysis?

Sometimes one set of information that includes all groups will hide the fact that there is a great discrepancy in that issue between some of those groups. It may be important to collect information on an indicator separately for each group. This is called data disaggregation. Typically, information is collected in ways that reflect the components most relevant to the project such as gender, identity, ethnicity, age, or area of origin. For example, if we are considering youth involvement in politics, an increase in the number of youth joining political parties may conceal the fact that young women are not becoming more involved. See page 216 in the Methods chapter for more information on data disaggregation.

What are the risks in working with indicators?

“When you’re up to your eyeballs in crocodiles, it’s hard to remember you were trying to drain the swamp.”

- Unknown

Humans want to succeed, particularly when being evaluated. As programs develop more reliable and valid indicators, there is a temptation to replace objectives with indicators. Some of the best-known examples of this phenomenon are in the education field where test scores were previously used as indicators of learning. As test scores become increasingly important for college education or school performance reviews, test performance has superseded and replaced the original learning objective. What used to be, “Achieve basic competency in algebra” has become, “Achieve at least 650 on the math portion of the aptitude test.” The result is improved test scores, but little change in the targeted skill and no evidence of skill use. The primary skill that improved is test-taking rather than skills in algebra.

A similar situation can be found in mediation programs that use the percentage of mediations resulting in written agreements as an indicator of success. It becomes part of the culture. At the end of mediations, program staff members ask mediators, “Did the parties reach a written agreement?” This conveys the message that this type of outcome is the most desirable. Parties who elect to conclude a mediation without a written agreement often believe that the mediation process was a success. Is the purpose of mediation to produce written agreements or is it to provide a safe forum for communication and dispute resolution? As Einstein said, “Not everything that can be measured counts.”

How do we develop indicators?

Many fields, such as health, political science and others, have already invested in research to develop indicators. This is often the work of professional researchers and social scientists. The challenge comes in developing units of measure for intangible issues such as trust, acceptance, and reconciliation. That challenge is made even more complex because of the importance of local context, culture, and perception. To make it still more difficult, we tend to put off thinking about indicators until we're late in the planning stages and facing tight deadlines.

Despite these challenges, peacebuilding practitioners can and need to develop new indicators and build on existing ones. Good indicators are context specific. Practitioners are supposedly well-immersed in the context and, hence, well-positioned to develop locally relevant indicators. Teaming up with social scientists may facilitate the process.

Here are a few tips for developing indicators:

1. Preparation

- Articulate your theory(ies) of change as well as the types of change specified in the objectives.
- Begin by determining what indicators have already been developed, tested, and refined in your program and other programs with which you are in contact. Keep an indicator bank. Ask other practitioners in peacebuilding for their indicators. Designate someone on your team as the indicator scout. Keep up-to-date with research on peacebuilding.
- Borrow from other professions. Public health is particularly rich in this area given its emphasis on behavioral change. Political science and psychology also offer a number of related indicators. Obviously, one cannot cut and paste indicators from other fields and expect them to work for peacebuilding. Nonetheless, indicators from other fields can serve to spark creativity in the development of indicators for peacebuilding.
- Make indicator development a continuous undertaking. Allocate time for learning and reflection. Unless you thrive under pressure, program design is not the best time to be developing indicators.

2. Generation

- Brainstorm all related things or dimensions that can be counted, measured, or sized and look for creative ways to combine some of those.

- Ask the parties or stakeholders in the conflict what they consider to be significant signals of change. When two women serving as community dispute resolution facilitators in Burundi were asked how we could prove that their work had resulted in change in their communities, they replied, “You could go talk to the local magistrate and ask him how his work has changed now that he refers so many cases to us.” Clearly, the local magistrate had come to value the dispute resolution work of these women.
- Break issues into smaller components. Rather than measure reconciliation, consider its components: mercy, justice, truth, and peace. To measure capacity, focus on skills, technical knowledge, process, motivation, and opportunity. This practice also goes by the name of factor analysis where all factors that influence the change are identified and, where possible, weighed according to the degree of influence each factor has.
- Use deductive logic. What would we have to see to know that objective X has been achieved? For example, seeing Hutus circulate in formerly Tutsi-only neighborhoods might mean that participants feel secure enough to expand their circle of mobility. Of course, we would need to verify that the Tutsis are still there.

Good indicators are context specific.

Map out the dynamics or factors involved. Common mapping methods include systems mapping, factor analysis, and force field analysis. Systems mapping can be helpful in illustrating and understanding how certain elements in a system can have far-reaching consequences if changed. For more information on systems mapping, see *The Fifth Discipline* by Peter Senge.

Advanced Concept

Borrowing from Social Capital

In a landmark study on social capital for the World Bank, Narayan and Cassidy mapped out their understanding or hypothesis of the factors influencing social capital as a framework for their research. They then looked for indicators of change in those factors. Their approach illustrates one way of developing indicators and some of the challenges and difficulties in working with intangible issues.

Social capital refers to the resources that accrue to a person or group by virtue of their relationships and networks. The map on the following page illustrates the different factors that Narayan and Cassidy believe contribute to social capital. It also illustrates their thinking or logic on how different phenomena are interconnected. Their paper includes specific questions that serve to detect signals of changes within each of the factors. The authors link specific types of change (e.g., the behavior of asking a neighbor

to care for a sick child) that contribute to a larger factor (e.g., neighborhood connectedness), which when combined with other factors influence social capital. Every country and every culture requires its own factors (or at least factors adjusted and amended to reflect local realities).

The work on social capital is also cited here because of subsequent efforts to look at the relationship between social capital and violence. For more information, see The Nexus between Violent Conflict, Social Capital and Social Cohesion: Case Studies from Cambodia and Rwanda by Nat J. Colletta and Michelle L. Cullen (World Bank, 2002).

3. Refinement

- Keep focused with the mantra, “What do I need to know and what information will tell me what I need to know?”
- If your objective is not providing clarity, consider rewriting your objective.
- Once you have a couple of possible indicators, look for ways to make them increasingly simple.

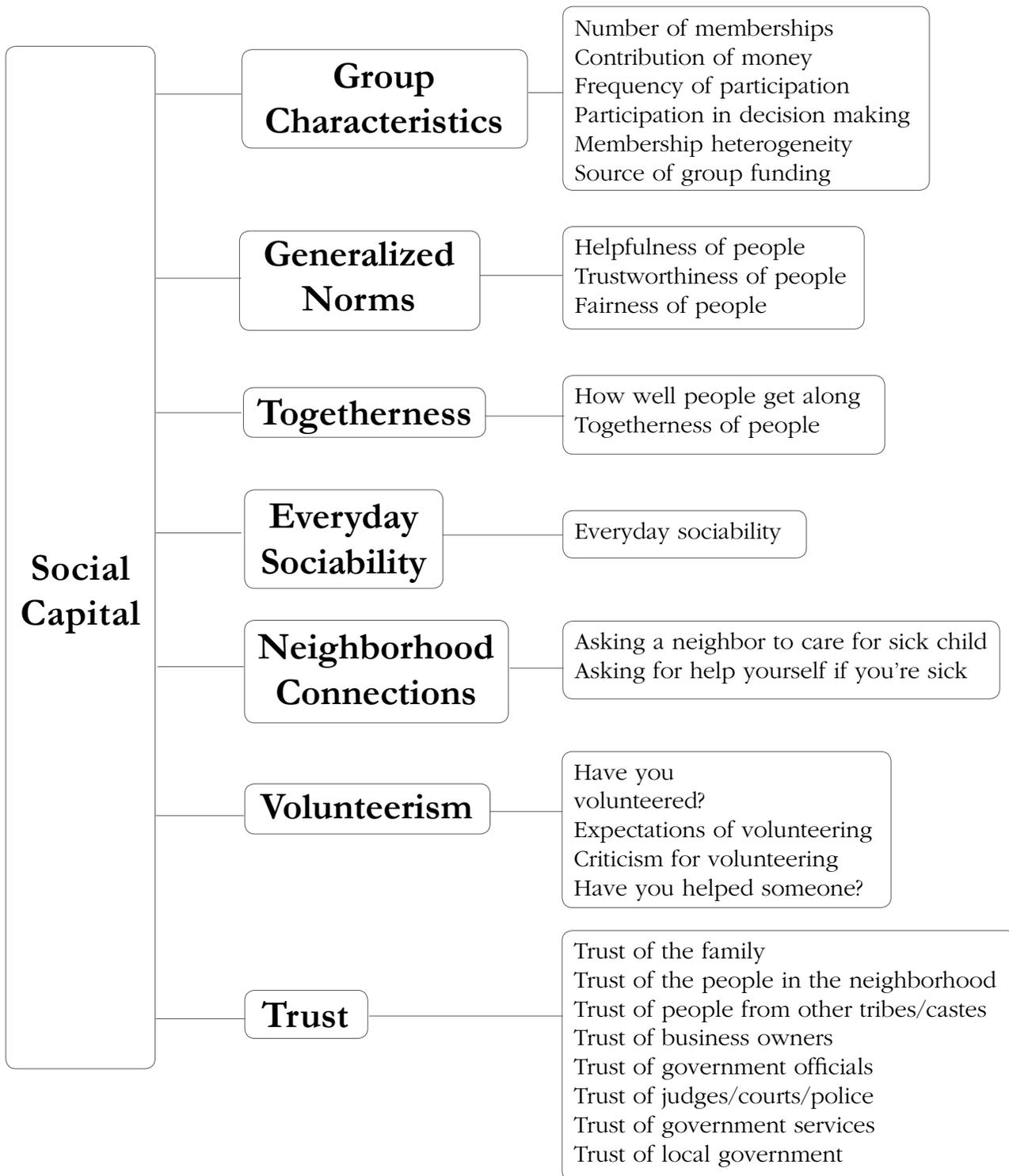
4. Testing Indicators

Many people invest great amounts of time and energy developing new indicators and then wait until the middle of the program — or worse, wait until the end — to put them to use. They risk an unpleasant surprise, and potentially their credibility with donors, if the indicators are flawed or not useful. Experience suggests that it is prudent to test new and newly modified indicators for their utility in decision making as early as possible, preferably during the design phase while there is still time to make changes. Test both the most promising indicators and those which are also viable but not as ideal. This helps ensure there is an alternative in the event that the ideal indicators don't survive the test.

This test is different from a test of the data collection method. The idea is to determine the utility of the indicators in the analysis and subsequent decisions. Pick hypothetical extremes using fictitious data and consider how the different extremes will influence decisions. If vastly different information has no influence over the decisions, the indicator is probably not useful and should be changed.

Experience suggests that it is prudent to test new and newly modified indicators for their utility in decision making as early as possible.

The Dimension of Social Capital, Narayan and Cassidy



If indicators tell us what has changed but not why, how do we find out why it changed?

The underlying assumption is that the change took place because of the activities the project implemented. However, in working with intangible changes, this assumption can be very difficult to substantiate. A more credible approach is to identify and acknowledge all the major contributing factors and illustrate how the program contributed along with the other factors. This implies some analysis and understanding of the many forces bearing on the issues under consideration.

Consider the rapid increase in human rights abuses by the military in Burundi in the late 1990s. The work of Human Rights Watch and others helped substantiate the increasing abuses. Why was this change happening? The factors contributing to this increase in abuses included massive and rapid recruitment, an increase in the ratio of soldiers to officers, a reduction in the duration and quality of basic training, poor supervision, and an atmosphere of impunity. Indicators can tell us that change is occurring, but determining why change happens requires additional research and analysis

Advanced Concept

There are a number of other more sophisticated types of indicators outlined in the following table. The examples illustrate how each type of indicator might fit within the example's intended outcome.

Goal: Increase acceptance of interdependence of all parties to the conflict.

Intended Outcome: Increased knowledge of sensible, responsible people on all sides of the conflict.

Potential Indicators:

Examples of Advanced Indicators¹⁰

| Type of Indicator | Description | Example Indicator | Performance Question |
|----------------------------|--|--|--------------------------------------|
| <i>Simple Quantitative</i> | <i>Requires only one unit of measurement (in addition to time)</i> | <i>Number of neighborhoods using program-related processes</i> | <i>Are we covering enough areas?</i> |

¹⁰ Adapted from *Managing for Impact in Rural Development, A Guide for Project M&E*, Section 5-21, IFAD 2000.

Examples of Advanced Indicators

| Type of Indicator | Description | Example Indicator | Performance Question |
|-------------------------------|--|--|---|
| <i>Complex Quantitative</i> | <i>Combines two or more units of measure (in addition to time)</i> | <i>Number of respected leaders who maintain at least three new relationships with people from the other side during the first six months of the project</i> <i>Percentage of people who can name at least three people from the other side who they consider sensible and responsible at the end of 10 months</i> | <i>Are enough respected people with influence engaged in preparatory work?</i> <i>Are new relationships being established?</i> |
| <i>Compound</i> | <i>Contains some sort of standard that requires definition or additional assessment</i> | <i>Spheres of influence of participating respected leaders, mapped quarterly</i> <i>Or</i> <i>Increase in capacity to communicate with the other parties, assessed quarterly</i> | <i>Are we reaching enough people? Where are the gaps?</i> <i>Or</i> <i>Do people have the skills, knowledge, resources, and motivation needed to communicate?</i> |
| <i>Scales and Indices</i> | <i>Scales or indices combine multiple indicators. Relatively rare in peacebuilding and conflict resolution</i> | <i>Change in ranking on the Awareness Scale/Index</i> | <i>How and how much has the awareness of the others' interests improved?</i> |
| <i>Proxy Indicators</i> | <i>A symbolic or approximate change relating to the desired outcome</i> | <i>Ratio of the use of non-judgmental language versus the use of judgmental language in intra-group gatherings during the first six months of the project</i> | <i>Are people talking about the conflict among themselves in new and constructive ways?</i> |
| <i>Open-ended Qualitative</i> | <i>Allows respondents to determine the qualities of the project that they deem to be important</i> | <i>Perceptions of the people about the accomplishments of the project</i> | <i>How does the larger community perceive the project?</i> |
| <i>Focused Qualitative</i> | <i>Focused on specific qualities of interest</i> | <i>Percentage of respondents who perceive positive changes in relationships at the end of year one of the project</i> | <i>How does the community perceive changes in relationships?</i> |



“We neglected to do a baseline at the beginning of the conflict. However, we just got an accurate carbon 14 reading on their positions!”

Written by *M. M. Rogers* and illustrated by *Ariv R. Faizal, Wahyu S., Ary W.S.*
Creative team for Search for Common Ground in Indonesia

What are the pitfalls to universal conflict transformation indicators?

Local realities and unique contexts make universal indicators difficult and somewhat risky to use.

Numerous attempts have been made at developing universal and generic indicators. It would certainly be helpful to be able to choose indicators off the shelf. However, local realities and unique contexts make universal indicators difficult and somewhat risky to use. The social capital example in the Advanced Concept on page 57 illustrates the great lengths to which researchers have gone to adapt indicators to each of the cultures and situations they were considering. This is not to say that new indicators must be developed from scratch for every program. Rather, they need to be modified and made contextually relevant. This may be easier than developing new indicators from a blank slate.

In sum...

Indicators are a means to measure change. They are the raw materials for much of monitoring and evaluation. Yet indicators are approximations, based largely on assumptions: the smaller or more accurate the assumption, the more reliable the indicator. A mix of qualitative and quantitative indicators usually reveals nuances and greater insight into what is happening. Given the difficulty in creating peacebuilding indicators, once indicators have been found, there is a risk of losing sight of the objective by over-emphasizing the indicator. Indicators borrowed from other fields or other cultures always need to be reconstituted and tested for each particular context and culture.

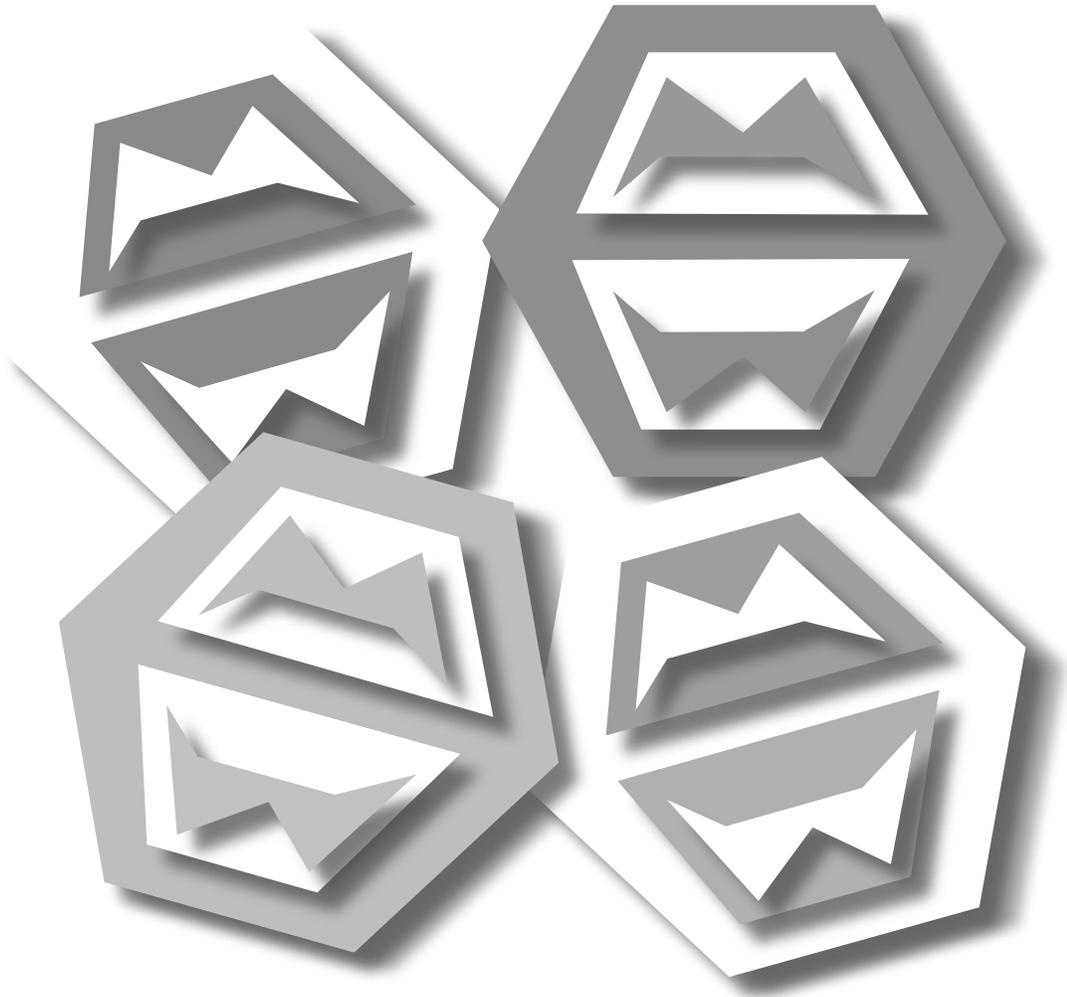
Further Reading:

Nat Colletta and Michelle Cullen, *The Nexus between Violent Conflict, Social Capital and Social Cohesion: Case Studies from Cambodia and Rwanda*, Social Capital Initiative Working Paper No. 23, World Bank, 2002.

Deepa Narayan and Michael F. Cassidy, *A Dimensional Approach to Measuring Social Capital: Development and Validation of a Social Capital Inventory*, Current Sociology Vol. 49 No. 2, Sage Publications, March 2001.

Michael Quinn Patton, *Utilization-Focused Evaluation*, SAGE Publications, 1996.

UNDP, RBM [Results-Based Management] in UNDP: *Selecting Indicators, Signposts of Development*, (UNDP, 2002).



BASELINE

This chapter includes:

1. Description of a baseline
2. Difference between conflict assessments and baselines
3. Baseline plans
4. Who conducts baselines and when

INTRODUCTION

“The term up has no meaning apart from the word down. The term fast has no meaning apart from the term slow. In addition, such terms have no meaning even when used together, except when confined to a very particular situation...”

- THURMAN W. ARNOLD

A baseline provides a starting point from which a comparison can be made.

Baselines are the most often forgotten component within design, monitoring and evaluation, yet they are key to proving that change has truly taken place. This chapter describes what constitutes a baseline and contrasts that description with conflict assessments. It then provides a planning tool for baseline development. Finally, it covers a number of the practical issues relevant to implementing baselines.

What is a baseline?

A baseline provides a starting point from which a comparison can be made. It is conducted prior to the beginning of the intervention and is the point of comparison for monitoring and evaluation data. The bulk of baseline studies focus on the intended outcomes of a project. They can also take into account secondary outcomes and assumptions, though these are not the primary emphasis.

What is the difference between a conflict assessment and a baseline?

A conflict assessment is an exploration of the realities of the conflict and an analysis of its underlying causes. An assessment can be done at any time, independently of a program or as a part of an existing program. Assessments are often conducted to determine whether an intervention is needed and, if so, what type of intervention. In a sense, an assessment is the basis from which the programming will be designed. Conversely, a baseline identifies the status of the targeted change before the project starts but after it has been designed.

Assessments and baselines should not be blended together. Nor should one be used as a substitute for the other since their *raison d'être*, focus, and implementation are very different.

Distinguishing Conflict Assessments and Baseline Studies

| | Conflict Assessment | Baseline Study |
|-------|--|---|
| Why | <ul style="list-style-type: none"> • Understand key factors and actors • Inform strategy | <ul style="list-style-type: none"> • Establish the status of the intended changes as a point of comparison |
| Who | <ul style="list-style-type: none"> • Staff, external consultants, or blend | <ul style="list-style-type: none"> • Ideally, this is the same person who will conduct the evaluation |
| When | <ul style="list-style-type: none"> • Before the project design | <ul style="list-style-type: none"> • After the design and before the implementation |
| Where | <ul style="list-style-type: none"> • Ideally in the conflict area, though desk-based is possible | <ul style="list-style-type: none"> • Conflict area |

Example

Land Claims Conflict

Part 1: Consider a conflict environment where the conflict analysis shows:

- Misinformation around land claims is a significant cause of violence
- Lack of access to official land registry offices make it difficult for legitimate landowners to get the appropriate documentation
- District councils appear to be central actors in resolving land claim disputes
- Approximately 80% of district councils have land claim policies
- Approximately 90% of people interviewed did not know that the district council had a policy, although the majority of this grouping thought that it was the role of the district council to do something

Based on this conflict assessment, an intervention was designed to decrease violence initiated by land disputes. It has two objectives. The first is, “Land-owning population has increased knowledge of local government initiatives relating to land claims.” One of the indicators developed for this objective is “% of district councils which have implemented communication strategies on land claim policies in 1 year.” A more direct indicator for this objective would be “% of people who can correctly state land claim policies.” However, the project team felt that collecting data that represented the entire country for this indicator was not feasible.

The second objective is, “Increased percentage of land owners who have proper documentation regarding their ownership by participating in the district council land claims procedure.” The activities around this objective will start in July, six months after the activities for the first objec-

Before the activities for the first objective get started, the project team needs to know how many district councils are already implementing communication strategies on land claims. This will allow them to set the target within their indicator and to understand the difference their project is making since they will be able to draw a before-and-after comparison.

How is a baseline utilized?

Baseline information is used in a number of different ways:

- **COMPARE BASELINE INFORMATION WITH SUBSEQUENT INFORMATION TO SHOW THE CHANGE THAT HAS TAKEN PLACE OVER TIME.** For instance, a baseline conducted in 2003 in Angola showed that, before a project started, 75% of soldiers thought it acceptable to use violence against civilians. After two years of work consisting of trainings, a media campaign targeted at the public, and a knowledge-raising campaign within the rank-and-file of the military, a formative evaluation was conducted. The 2005 evaluation showed that 55% of soldiers think it is acceptable to use violence against civilians – a decrease of 20%.
- **REFINE PROGRAMMING DECISIONS ABOUT KEY STAKEHOLDERS OR POSSIBLE RESISTERS.** For example, an organization conducted a broad conflict assessment which determined that male youth ages 13-21 were the key recruits for two rebel groups in a conflict. It then created a prevention program focused on boys 10-13 years old. The baseline revealed that boys 15 and younger are less likely to be recruited; thus, there was a difference in age between those targeted by the program and those targeted by the rebels. In this case, prevention program target group could be refined to focus on the 12-14 year-old age group rather than the younger audience that was the original emphasis.
- **SET ACHIEVABLE AND REALISTIC TARGETS.** Consider the Angola military example once again. When the baseline study showed that 75% of soldiers thought it acceptable to use violence against civilians, the project team then knew the exact extent of the problem they wished to change. As a result, the team discussed the degree of change that would be realistic for them to expect to achieve based on the resources available for the project. At that point, they set the target for their objective: “At the end of a four-year project, less than 10% of the military will think it is acceptable to use violence against civilians.”
- **ENABLES MONITORING DATA TO HAVE GREATER UTILITY EARLIER IN THE PROJECT CYCLE.** Continuing with the Angolan military example, the project team gathered monitoring

data after the first six months of project implementation. They were able to compare it against the baseline to see if any progress had been made. The first time they collected data, it showed an increase in acceptance of the use of violence against civilians among soldiers rather than a decrease. The team discussed this finding and decided to hold the course until the next set of monitoring information was collected because they had actually only been working with soldiers for three months; the first three months had been spent in preparation. They hypothesized that the increase was due to more people hearing about the topic but not having enough time to internalize the actual message. Sure enough, six months later, the monitoring data showed a small but hopeful decrease.

- **ENSURES ACCURACY AND UTILITY OF INDICATORS.** The process of conducting the baseline itself provides the “testing” of the proposed indicators. This test informs the project team whether the indicators are an accurate signal and whether they can be measured. Consider a project with a goal of increasing cooperation between rural Hutus and Tutsis. One of the indicators proposed is, “An increase in neighbors of different ethnicity assisting each other in the fields.” However, when the baseline team went out to gather information, they discovered that, in the target communities, Hutu and Tutsi fields were rarely close together making it difficult for members from each community to assist each other even if they had wanted to. As a result, a new indicator was created.

Once the baseline data is available, the project team should be convened to review the results. This review is not an abstract discussion. It should include setting the targets for objectives and indicators. A discussion on whether the baseline information suggests that any element of the project needs to be refined should also occur. The project team should leave this meeting with a common understanding of the starting point for the project.

The baseline report and raw data should be stored so that the evaluation team can verify the conclusions, and/or analyze the raw data from a different perspective. See page 201 of the Methods chapter for more information.

Example

Land Claims Conflict

Part 2: The baseline study showed that 30% of district councils had communication strategies. However, only 10% of the communication strategies included the land claims policy components. The project team members can use this information to set their indicator targets as well as to refine the work plan.

The team determined that their one-year target would be that 75% of district councils would have communication strategies with a land claims component.

With this information, the project team can now refine their activities targeting those district councils with communication strategies differently than those councils who have no strategy at all.

What does a baseline focus on?

Baselines explore three areas:

- Change (outcomes)
- Secondary changes
- Assumptions

The first area, change, is required of all baselines, while the latter two, secondary changes and assumptions, are optional depending on the project. The majority of baselines focus on collecting data on the status of the targeted change (outcome or outcome-level indicators) before the project begins. For outcomes that do not require indicators, data is collected for the outcome (change) itself.

It is feasible, though not recommended, for impact-level change to be the focus of a baseline. Since the impact a project will have often requires a longer timeframe (5-10 years) to occur, focusing the baseline on a project's impact may result in selling the project short if its duration is not long enough.

Consider a project whose goal is to change public opinion in the Basque region of Spain from acceptance of political kidnappings as an appropriate tactic to non-acceptance of such acts. The project received funds for one year. The baseline gathered data on indicators of goal achievement only. After one year, many changes had occurred as a result of this project, such as an increase in the public's knowledge of the frequency of kidnappings and a significant change in the tone of media articles portraying the incidents. However, little movement had taken place at the public opinion level. The similar and more immediate changes (outcomes) were not captured in the evaluation because it only looked at impact (long term) achievement. If the baseline had focused on outcomes, the evaluation would have shown significantly better results. A compromise between the two is to select a few impact indicators that are deemed most important and include them in the baseline study. These indicators can provide useful information on an overall shift.

The second area that a baseline can include is secondary changes, both positive and negative. If a project team wants to understand the effects

on or relationship between their project and an indirect target, this is a secondary change. The direct change is not what is being examined by looking at secondary changes; rather, there is a belief that the project will affect an indirect target in some way. Of course, this effect could be positive or negative. Gathering data on the effects of the intervention in terms of a secondary change can be very useful in advancing the field's understanding of how peacebuilding projects affect the environments in which they operate.

Consider a television project in Macedonia directed at children ages 6-11 that challenges negative stereotypes of the "other." In addition to the direct impact of the show on the children watching, the project staff members also wanted to know if children can influence their parents' thinking about the "other." If so, the parents may be an indirect target of this intervention; therefore, the project team would like to understand more about the secondary changes to this indirect target. To elicit such information, project team members added parents to the list of people to be interviewed during the baseline about their attitudes and behaviors on ethnic relations.

Assumptions about the objectives also need to be considered in determining what information to obtain in the baseline. The question is, what information will help program managers determine if the assumption continues to hold true? Consider the example of a project that seeks to reduce incidents of youth violence. One of the activities within this project is to introduce community policing in particularly violent neighborhoods. The assumption associated with this activity is that neighborhoods will participate in community policing. Useful information to gather in the baseline would be neighborhood knowledge and attitudes about community policing. This would not only inform the strategy, it could also be tracked over time as part of the indirect changes enacted by the project. More on this example may be found in the Design chapter on page 25.

A baseline plan illustrates what information is needed as well as how, where, and from whom it will be collected. It is very similar to an evaluation plan, a description of which can be found in the Evaluation Management chapter on page 153.

Example

Land Claims Conflict

Part 3: The project team understood that "change" had to be a focus of their baseline study. As a result, indicators for each objective were included. The results for one of those indicators showed that only 10% of district councils had communication strategies with land claim components included.

Example

Because the organization had run this type of project in many other countries, the team manager felt that the model was well-developed. She had read the previous project designs and evaluations as well as the lessons learned documents that were generated about this approach. Consequently, she felt that there were no secondary outcomes to be added to the baseline.

In terms of assumptions as a baseline focus, there is some additional information to gather. The first objective is “land-owning population has increased knowledge of local government initiatives relating to land claims.” The assumption is that the government initiatives to address land claims are effective responses to the problems. In the baseline, data on people’s perceptions of the effectiveness of the governmental land claim policies was collected.

A baseline plan illustrates what information is needed as well as how, where, and from whom it will be collected.



Early efforts at certifying evaluators

Written by *M. M. Rogers* and illustrated by *Lawson Sworb*

What is a baseline plan?

The table below shows the different areas that should be included in a baseline plan. The baseline focus column always contains the intended

change (outcomes) and may include optional areas such as secondary changes or assumptions. The optional areas are depicted by the asterisk in the diagram.

Baseline Plan

| Baseline Focus | Indicators or Line of Inquiry | Means of Verification (MOV) | Data Source & Target | Location of Data Collection | Conflict Considerations | Means of Analysis | Time Needed |
|-------------------|-------------------------------|-----------------------------|----------------------|-----------------------------|-------------------------|-------------------|-------------|
| Change (Outcomes) | | | | | | | |
| Secondary Change* | | | | | | | |
| Assumptions* | | | | | | | |

What does a baseline plan contain?

BASELINE FOCUS: There are three possible areas of focus: change, secondary outcomes and assumptions. Descriptions of each of these areas may be found on page 66 earlier in this chapter.

INDICATORS OR LINES OF INQUIRY: Pending the baseline focus, indicators or lines of inquiry may need to be developed to further direct the study. A line of inquiry is a more specific question or set of questions than the focus. An indicator, on the other hand, is used where an outcome cannot be measured directly. See page 142 in Evaluation Management for more information on lines of inquiry.

For the change focus of the baseline, the indicators for each outcome should be found in the design tool, which are most often logical frameworks or results frameworks. Alterations to the indicators can be made at this time; in fact, there is no better time to refine them. The project team needs to be fully supportive of any refinements to, or additions of, indicators since it is best to not change them again after this point. Altering the indicators later would probably make the baseline useless unless the new indicator is based on some reconfiguration of the same data. Of course, if the situation on the ground has changed drastically by the time the evaluation occurs, the indicators may nevertheless need to be altered in order for them to be useful in the new reality.

Indicators are also needed for secondary outcomes. Indicators for secondary outcomes would generally be developed when the baseline plan is created since they are not normally part of a design tool. Finally, when assumptions are included in a baseline, they frequently require lines of inquiry to be developed for each assumption so that the evaluator has a more specific question to gather data for. An example is found on page 71 later in the chapter.

MEANS OF VERIFICATION (MOV): The MOV, also called data collection methods, is the way in which data will be collected. The methods available for evaluations are also available for baseline studies. Fundamentally, method selection is driven by the information one is trying to find. Where possible, use the same methods for the baseline and the evaluation. This saves time and money since new instruments or tools need not be developed and tested. Assuming that quality instruments were developed for the baseline, use of the same methods also enhances the accuracy of the results. However, if there are drastic shifts in the context or perceived flaws in the baseline approach, new methods may need to be selected for the evaluation. More on data collection may be found in the Methods chapter, page 201.

Methods selection is based on the best way to access the information being sought. Different data collection methods may be utilized for each baseline focus. Similarly, different methods may be chosen to gather data on the same focus. On the other hand, the same method can be used for each focus. For the change focus, the MOV is generally a required part of the design tool. The methods to collect data on the secondary outcomes and assumptions will need to be developed with the baseline plan because these are generally not part of the average project design process.

DATA SOURCE AND QUANTITY: The data source and target refer to where the data will be accessed and how many data sources will be utilized. For instance, a data source and target might be 80% of the participants in training.

LOCATION OF DATA COLLECTION: Where will the data be collected? Using the training example from the previous paragraph, will it be during the training, in their homes, electronically via e-mail or a website, or in their place of work?

CONFLICT CONSIDERATIONS: This includes issues specific to the conflict that may affect the baseline such as security environment, implication of language selection, nationality of researcher, or avoiding contentious memorial days in the conflict. See Evaluation Management, page 137, for a more in-depth discussion of this issue.

TIME NEEDED: Time depicts the number of days to implement each aspect of the baseline including doing the analysis of the data.

A line of inquiry is a more specific question or set of questions than the focus.

Baseline Plan Example

| Baseline Focus | Indicators or Line of Inquiry | Means of Verification (MOV) | Data Source & Quantity | Location of Data Collection | Conflict Considerations | Means of Analysis | Time Needed |
|---|---|--|---|-----------------------------------|--|--|-------------|
| Change | | | | | | | |
| Objective 1: Reduce negative stereotypes of the “other” held by teenagers | % increase in number of children who have friends of the other religion [indicator] | Direct observation | Approximately 40 children from 4 “hang-out” areas accessible by both communities* | Play areas throughout the village | | Review of observation notes | 1.5 days |
| Objective 1: Reduce negative stereotypes of the “other” held by teenagers | % decrease in the number of children who associate negative images with descriptions of the “other” [indicator] | One-on-one interviews | 6 teachers per school (3 schools) | In a private office of the school | Nationality of the interviewer Language of interview | Review of interview transcripts utilizing pre-set criteria | 3.5 days |
| Objective 1: Reduce negative stereotypes of the “other” held by teenagers | % decrease in the number of children who associate negative images with descriptions of the “other” [indicator] | Photographs of what represents the “other” | 50 teenagers* [25 teenagers from each religion] | Village | Security concern of those taking photos. Will it be acceptable or cause offense? | Photos will be categorized according to types of images | 1.5 days |
| Objective 2: Equip teenagers with the skills and motivation to use conflict resolution techniques when conflict arises | What skills do students currently utilize in dealing with conflicts? [line of inquiry] | 4 focus groups | 32 parents* [16 parents from each religion] | Held in the schools | Identify appropriate language and sensitive words for questions | Review of transcripts | 4.5 days |
| Objective 2: Equip teenagers with the skills and motivation to use conflict resolution techniques when conflict arises | Role plays with discussion groups | Role plays with discussion groups | 20 teenagers* [10 teenagers from each religion] | Held in the schools | Identify appropriate language and sensitive words for questions | Assess role plays against specified behaviors and language | 2 days |

Baseline Plan Example

| Baseline Focus | Indicators or Line of Inquiry | Means of Verification (MOV) | Data Source & Quantity | Location of Data Collection | Conflict Considerations | Means of Analysis | Time Needed |
|---|---|--|--|--|--|---|-----------------------|
| Objective 2: Equip teenagers with the skills and motivation to use conflict resolution techniques when conflict arises | % increase in number of conflicts that teenagers use conflict resolution skills [indicator] | One-on-one interviews Focus groups | 6 teachers from each school 32 parents* [16 parents from each religion] | In a private office of the school Held in the schools | Language of interview Nationality of the interviewer Identify appropriate language and sensitive words for questions | Review of interview transcripts utilizing pre-set criteria Review of transcripts | .5 day .5 day |
| Hypothesized Outcomes | | | | | | | |
| Conflict resolution skills utilized in family settings | % decrease in domestic violence | Secondary data review | Official police statistics | Central police station in village | Consider the gender makeup of the police Consider the domestic violence reporting system | Document review | 1 day |
| Assumptions | | | | | | | |
| Negative stereotypes foster violence against the “other” | What percentage of the causes of violence between teenagers of different religions is not the result of negative stereotypes? [line of inquiry] | Photographs of causes of violence One-on-one interviews | 50 teenagers* [25 teenagers from each religion] 10 local police officers who work in different areas of the village | Village In their offices | Security concern of those taking photos. Will it be acceptable or cause offense? | Photos will be categorized according to types of images Review of notes | 1 day 3.5 days |

* Indicates that the data will be disaggregated by gender.

Baseline Plan Example

Consider a project that is being developed to target children ages 12-15 in a large village in Central Sulawesi, Indonesia. The goal of the project is to decrease “tension” between Christian and Muslim youth age 12-15 in the village. The project has two objectives and has just received confirmation from the donor that it will be funded. Refer to Baseline Plan Example chart on page 71.

PROJECT LOGIC:

Goal: Decrease tension between Christian and Muslim youth ages 12-15 in the village.

Objective #1: Reduce negative stereotypes of the “other” held by teenagers ages 12-15 within the village.

Assumption: Negative stereotypes foster tension against the “other.”

Objective #2: Equip teenagers ages 12-15 within the village with the skills and motivation to use conflict resolution techniques when conflict arises.

Activities: The peacebuilding work includes a comic book series based on the reality of Central Sulawesi. The characters in the comic books model conflict resolution techniques and positive images of the “other.” In addition, there will be a curriculum for teachers to use in schools based on the comic book series. The teachers will also be trained in how to teach and discuss the issues in the curriculum.

Definition: The term “tension” is defined very broadly to include bullying, swearing, graffiti, and all other forms of intimidation.

The project team wanted data on both objectives (i.e., reduction of negative stereotypes and providing conflict resolution skills), one secondary outcome, and one of their core assumptions. A variety of methods were used, many applying to more than one focus area.

The indicators had been well-developed by the project team so they required no refinement by the evaluators. The indicators included:

Objective 1: % increase in number of teenagers (ages 12-15 within the village) who have friends of the other religion

Objective 1: % decrease in the number of teenagers (ages 12-15 within the village) who associate negative images with descriptions of the “other”

Objective 2: % increase in the number of conflicts in which teenagers (ages 12-15 within the village) use conflict resolution techniques when conflict arises

The first objective, “Reduce negative stereotypes of the “other” held by teenagers,” has two indicators, both included in the baseline. The data for the first indicator, “% increase in number of teenagers who have friends of the other religion,” will be collected using two means: direct observation and one-on-one interviews with teachers. Two data collection methods were selected in order to strengthen the validity of the information gathered.

The direct observation (method) will be conducted at four different hang-out areas (location of data collection) throughout the village where it is known that teenagers of both religions frequent but rarely mix. It is hoped that approximately 50 teenagers (data source and quantity) will be observed through this method. Every effort will be made to ensure that there are boys and girls represented in these areas. A local parent in each area will assist the evaluator in determining the religion of each teenager.

Prior to collecting the data, the evaluator will develop an observation guide that outlines what she/he is looking for, such as terms used in conversations or types of behavior. The data will be broken out by gender as well (means of analysis). Developing the observation guide, gathering data, and analyzing the results should take approximately 1.5 days (time).

To supplement the observation the evaluator will also do one-on-one, semi-structured interviews (method) with six teachers (data source and quantity) from each of the three schools in the village for a total of 18 teachers. The teachers will be from different age groupings in the school. The interviews will be conducted in a private office within the school (location of data collection) to ensure that the teachers feel comfortable speaking truthfully. In developing the interview guidelines, the evaluator will need to ensure that the language selected for the questions is appropriate, unbiased, and non-inflammatory, and avoids taboo topics. She/he should also check with the project team to see if language appropriateness differs between the two communities. In addition, the evaluator should investigate the potential effects and perceptions that people may have due to her/his nationality.

The notes from the interviews will be analyzed against pre-set criteria (means of analysis). Developing the interview guide, arranging the interviews, collecting data, and doing the analysis will take approximately 3.5 days (time).

Data for the second indicator, “% decrease in the number of teenagers who associate negative images with descriptions of the ‘other,’” will

be gathered through photographs (methods) taken by 40 teenagers (data source and quantity) depicting what they feel are images of the “other.” Half of the group will be Christians and the other half Muslims, and they will be from different areas of the village. The evaluator will need to take care that the act of taking pictures will not be seen as provocative or offensive (conflict considerations). The photographs will then be categorized (means of analysis). The entire process will take approximately 1.5 days (time).

The second objective is, “Equip teenagers with the skills and motivation to use conflict resolution techniques when conflict arises.” The project team wanted the baseline to investigate one line of inquiry and one indicator. The team wanted more information on the skills students currently utilize in dealing with conflicts (line of inquiry). To gather this data, four (data quantity) focus groups (method) of parents (data source) will be conducted. Two groups will be of the Christian religion and two groups Muslim (conflict consideration). The sessions will be held in the appropriate local school (location of data collection). Care needs to be taken in developing the language of the questions for the facilitation (conflict consideration). The results will be analyzed through a review of the transcripts (means of analysis), all of which should take approximately 4.5 days (time).

The other data method will be role-plays done by 20 teenagers (data source and quantity). The teenagers will need to have gender and religious equity in their numbers (conflict consideration). As time is limited, the teenagers who will be taking photographs will be invited to provide the names of other teenagers who could be invited to the role-plays. The role-plays will depict common conflict situations and the teenagers will need to behave the way they think someone from the “other” community would behave in that situation. Immediately following the role plays, small group discussions will be held to review how the roles were depicted and if that was realistic to life situations. Held in schools (location of data collection), the evaluator will assess the plays and the subsequent discussions against a specified set of behaviors and language (means of analysis). It is estimated that this will take two days to complete (time).

The second objective also has an indicator, “% increase in the number of conflicts in which teenagers use conflict resolution skills.” Data collection for this indicator will be interwoven into methods previously mentioned: the one-on-one interviews of teachers and the focus groups of parents. It is anticipated that the additional instrument development and analysis will require an extra half-day for each method.

The project team believes that this intervention could also have some other indirect effects or secondary outcomes. Due to time and budget constraints, they selected one of these to investigate further: the use of conflict resolution skills in family settings. The team believes that, if

teenagers develop conflict resolution skills, they will not only apply those skills to conflicts with the “other” but they will also use them among friends and family to a positive end. To collect data on this potential indirect effect, an indicator was developed to look at the percentage decrease in domestic violence. A secondary data review (method) of official police statistics over the past six months (data source and quantity) will be conducted. The data will be sourced from the central police station in the village (location of data collection).

Consideration of the gender makeup of the police force, as well as the domestic violence reporting procedure, should also be part of the analysis. This effort should take approximately one day (time).

Finally, when the project team reviewed their assumptions, there was one for which they anticipated that further data collection would be beneficial: the assumption that negative stereotypes foster violence against the “other.” The line of inquiry chosen was, “In what amount are other reasons the cause of violence between teenagers of different religions?” Data on causes of violence will come from the photographs (methods) taken by teenagers as described earlier. The additional analysis should take approximately one day (time).

The information on the amount that those other causes spark violence will come from one-on-one interviews (methods) with ten local police (data source and quantity), conducted in their offices (data collection location). The work will take around 3.5 days to complete (time).

The evaluator conducting the study most commonly develops the baseline plan in conjunction with the project team.

Who develops the baseline plan and when?

The evaluator conducting the study most commonly develops the baseline plan in conjunction with the project team. Key elements of the plan, such as indicators, are generally derived from the project documentation. If they have the experience, the project team members can also develop the baseline plan, although they should not finalize it until the evaluator has been hired and can provide input.

If feasible, the baseline plan should be developed immediately prior to its implementation. This ensures that it reflects the most current situation on the ground as well as any changes in thinking by the project team.

Example

Land Claims Conflict

Part 4: The organization has a Design, Monitoring, and Evaluation (DM&E) Specialist as part of its regional team whose members have been supporting the intervention design. With the DM&E Specialist’s help, the project team determined what would be included in the baseline focus during one of the project meetings to finalize the design.

Example

Rather than develop the full baseline plan, the team contracted an evaluator who had delivered a high quality evaluation on a similar project in the region. (This sole-source approach for recruiting is explained further in *Managing Evaluations*, page 153). They decided that the evaluator should develop the baseline plan since this would be a more time-effective process. The evaluator drafted the remaining aspects of the baseline plan after many discussions with the project team. The team then offered their input and the evaluator reworked the plan accordingly.

The baseline occurs a few weeks before the intervention is implemented.

For complex programs where there are several different changes projected, each resulting from activities that start at significantly different times, a rolling baseline should be considered.

When does a baseline study take place?

The baseline occurs a few weeks before the intervention is implemented. Practically speaking, it is included as the first activity in the project work plan. The data gathered from the baseline then informs the target setting in the design tool.

This can be difficult to achieve in conflict contexts where it can be important to be on the ground very quickly or where sudden shifts can cause unexpected delays. When speed is of the essence, it is possible to conduct the baseline simultaneously with the first stages of the project. Conversely, there should not be a time lag between the baseline and the intervention start. The data should be collected reasonably close to the start of the project, with “reasonable” being defined by the rapidity of change in the context. A situation in rapid flux should attempt to minimize the time (e.g., 3-4 weeks) between the baseline and implementation, while a more stable situation could handle a longer delay (e.g., 2-3 months).

For complex programs where there are several different changes projected, each resulting from activities that start at significantly different times, a rolling baseline should be considered. In this approach, the baseline is broken into segments according to the different changes and is implemented prior to the start of activities for each new change. When using a rolling baseline for different implementation start dates, it is not simply a matter of different activities beginning; rather, the deciding factor is that the project is initiating a new change through the newly started activities.

Consider for example, a three-year project in Kosovo with a goal of decreased Serb-Albanian violence in the Peje/Pec municipality. The objectives include 1) increased community utilization of the UNMIK (United Nations Mission in Kosovo) war crimes investigation task force, 2) local truth and reconciliation structures operating within communities, and 3) increased trust in the Kosovo Police Service. Activities for the latter two objectives will start immediately after the receipt of funds, while the

activities for the first objective, the utilization of the UNMIK war crimes investigation task force, will start after 1.5 years. Before the project starts, baseline information should be gathered on the latter two objectives. At the 1.5 year mark of the project, the baseline should be conducted for the first objective.

A rolling baseline also applies for projects that have geographic rollouts. The project starts in one province and then moves to the next province after a set period of time (e.g., nine months). The baseline should be collected for each new provincial project launch. If there were 20 provinces and limited resources, one could select a portion of those provinces and conduct baselines in that portion only.

Example

Land Claims Conflict

Part 5: A rolling baseline will be used in such a manner that the second half of the baseline is implemented before the activities to initiate the second objective start. The first part of the baseline for this land claims conflict project was implemented within three weeks of receiving the project grant money. It gathered data on the first objective, “Land-owning population has increased information on local government initiatives relating to land claims,” because the activities to initiate this change were to start within eight weeks of receiving the funds. This timing, although tight, worked well because the data was brought back to the project team while there was still time to adjust the work plan and activities.

The second part of the baseline explored the second intended outcome, “Increased percentage of land owners who have proper documentation for their property.” The activities to initiate this change will occur nine months after the launch of the activities to increase the land-owning population’s knowledge of governmental land claims policies.

Who conducts the baseline?

The individual or team conducting the evaluation is the ideal choice to conduct the baseline. This approach ensures continuity between the baseline and evaluation and minimizes the chances of the baseline data being deemed incomplete or invalid. Moreover, it deepens the evaluators’ knowledge and implicit understanding of the context because they will have a picture of the “before” state. Of course, a different consultant who produces a high quality product may also be contracted, though this does not capitalize on the full package of potential benefits that would result from using the same people who conducted the evaluation and/or baselines.

An evaluation expert contracted to conduct a baseline can also provide valuable input into the project logic of the intervention. This expert can advise on the logic, indicator development, and the monitoring system as well. Since the project is not yet underway, it is a strategic time for this type of expert advice.

It is essential that the baseline be professionally implemented since it is pivotal to so many other steps in the project, from refining the design to measuring change. The methods must be appropriately selected and implemented well, and the analysis must be based on the evidence. A professional evaluation team will discard a baseline study that is not credible or that makes their job more difficult.

If contracting an external professional to conduct the full baseline is beyond the scope of the budget, and the specialized skills necessary do not exist on staff, consider hiring an external advisor. This person could still be a member of the future evaluation team, but would not implement the baseline in this scenario. Rather, she/he would advise on the methods selected, train the data collectors, and provide a quality check on the analysis.

We have no time to recruit an evaluator for the baseline, what do we do?

Baselines are generally not conducted unless funding has been secured, but once the funding becomes available, they should be rapidly implemented. This dynamic makes recruiting a qualified person challenging. If the time between receiving funds and the implementation start date is short, it is recommended to do the preliminary work for recruiting before the funds are awarded. This includes the development of the Terms of Reference and, potentially, the recruiting of an evaluator. Doing this advance preparation enables the baseline to be “on-the-ground” rapidly while still engaging qualified external support.

Example

Land Claims Conflict

Part 6: Since the baseline needed to be implemented very shortly after the funds were received, the team did some pre-planning for the recruitment of the evaluator. The team drafted the terms of reference, discussed evaluator options with the regional DM&E specialist, contacted the evaluator, and tentatively agreed on working terms. Upon receiving word that the grant was awarded, the team notified the evaluator so that dates could be blocked to conduct the baseline as soon as possible.

What should be done if a complete baseline is not possible?

If the timing and resources are such that a comprehensive baseline cannot be performed, there are two second-best options. The first is to conduct a condensed baseline where data is gathered on a few key indicators within the change focus. Start by selecting the key indicators either by identifying the ones that are most important overall or by selecting one for each objective.

The second option is to make the first monitoring exercise more comprehensive. This would involve collecting more data from more people than normally involved in a monitoring exercise. This is really a second best option since it does not provide the geographic or population coverage needed to make sound conclusions. Furthermore, because monitoring is generally conducted by staff members, the methods to be used will be limited to those that the staff members are competent in developing.

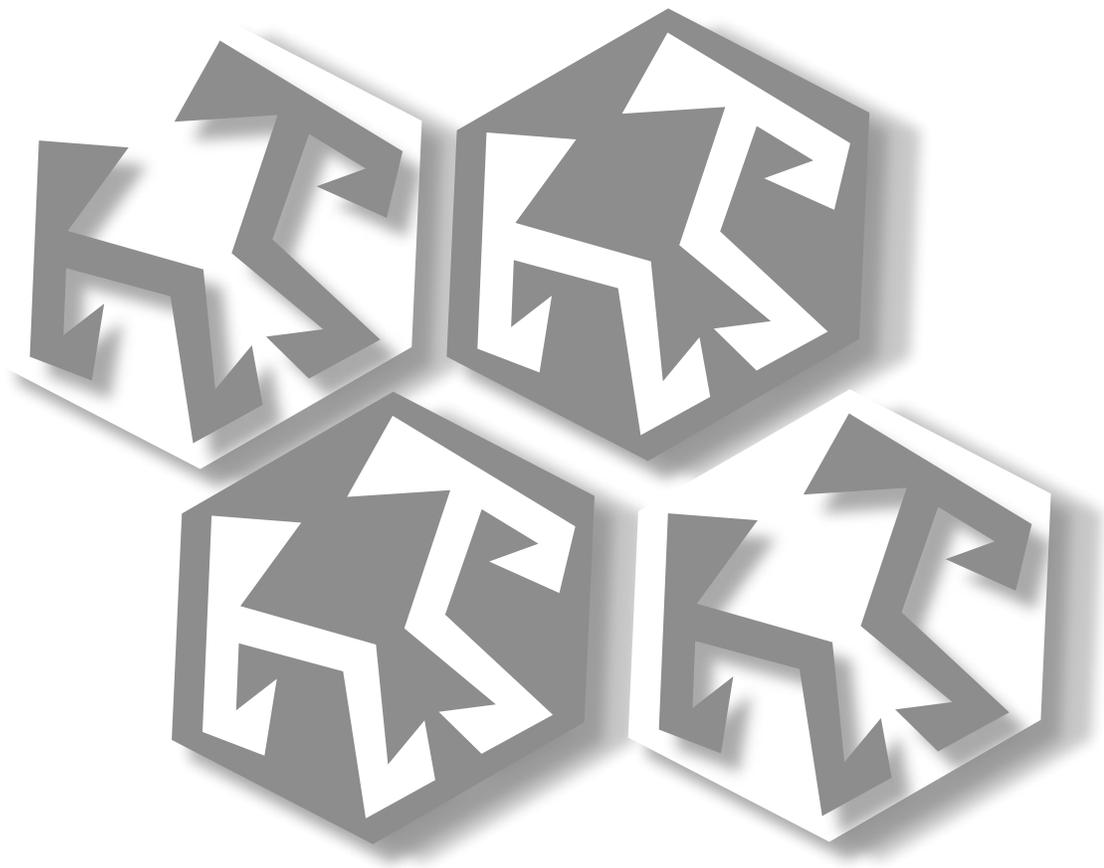
Not having a baseline does not invalidate the entire monitoring and evaluation process. Many evaluations are conducted in the absence of a baseline; however, the degree of conclusiveness of the findings is inherently limited where no baseline occurred. In other words, one can never say definitively that a change of X percent or from one status to another occurred since the beginning status is an approximation at best.

Many evaluations are conducted in the absence of a baseline; however, the degree of conclusiveness of the findings is inherently limited where no baseline occurred.

How do evaluators use baseline data?

Evaluators utilize baseline data as a point of comparison for the data that they collect during the evaluation. The baseline data should not be accepted as sound, however, without some form of verification. At a minimum, both the methodology and analysis should be reviewed. Doing a review does not mean redoing the entire analysis, but selecting a few key areas and checking to see if the approach was sound. Even when the person who conducted the baseline and evaluation are the same, a small portion of time should be taken to confirm that no mistakes were made in the baseline study.

For further information on recruiting consultants and managing evaluation projects, see *Evaluation Management*, page 137.



MONITORING

This chapter contains:

1. An explanation of the differences between monitoring and evaluation
2. Descriptions of three major areas for monitoring:
 - *Conflict and context*
 - *Implementation*
 - *Progress toward results*
3. An explanation of how monitoring is incorporated into design frameworks
4. An example of a monitoring plan

INTRODUCTION

“Where are we going?” asks the passenger as the vehicle races across the desert. “I don’t know,” replies the driver, “but we’re getting great gas mileage.”

- Excerpt from the film “Sahara.”

As illustrated in the above quote, much of the challenge in monitoring involves connecting relevant information to strategic decisions. This chapter discusses the relationship between monitoring and evaluation. It also covers the three basic types of monitoring in peacebuilding: the conflict context, program implementation, and progress toward results. Certain key or fundamental assumptions may require monitoring as well. In addition, the chapter briefly discusses reporting and explains how monitoring is integrated in the different design frameworks. An example of a monitoring plan is provided at the end.

Monitoring is an ongoing process that generates information to inform decisions about the program while it is being implemented.

How is monitoring different from evaluation?

Monitoring and evaluation are different sides of the same coin, which is but one of the coins in the currency of learning. Other coins in the currency of learning include action research, reflection, reading, coursework, literature research, and participatory rapid appraisals to mention only a few.

Monitoring is an ongoing process that generates information to inform decisions about the program while it is being implemented. Monitoring differs from evaluation primarily in terms of when and how often it is done and the decisions it informs. Generally, monitoring starts earlier and continues more frequently than evaluation. The decisions that monitoring informs are practical and detailed, and often meet an immediate pressing need or question.

Evaluation is more a multi-part event than a continuous process, and it often focuses on a bigger picture or on more complex issues such as why something happened. For example, in a program focused on the peace process, evaluation may look at overall advances in the peace process over time. Monitoring, in contrast, may focus on specific changes in communication channels, shifts in language on specific themes under negotiation, or changes in the number of alternatives under consideration by the parties.

Evaluation is more a multi-part event than a continuous process, and it often focuses on a bigger picture or on more complex issues such as why something happened.

Both monitoring and evaluation use data to inform decision-making and contribute to improved strategies. Both are also intended to generate lessons learned, although evaluation tends to contribute to more overarching lessons while monitoring contributes to more pragmatic or technical issues. Each of these disciplines demonstrates accountability. The following table illustrates some of the differences.

Distinguishing Monitoring and Evaluation

| | Monitoring | Evaluation |
|--------------------------|--|---|
| What is it? | Ongoing collection and analysis of data on progress toward results, changes in the context, strategies, and implementation | Reviewing what has happened and why, and determining relevance, efficiency, effectiveness, impact, etc. |
| Why do it? | Inform day-to-day decisionmaking Accountability and reporting | Strengthen future programming Provide evidence of success Deepen our understanding of how and why things work |
| Who does it? | Program Staff and/or Partners and/or Participants | External consultant, staff, participants or combination of these groups |
| When to plan | At design stage | Core decisions taken at design stage and refined prior to implementation |
| When to implement | Throughout the program – periodically, frequently or continuously | Mid-term (formative) Completion (summative) After completion (impact) |

Why develop a monitoring practice?

Monitoring involves the use of reliable data in timely and informed decision-making. Data and information are the foundations of a monitoring practice. Too often people confuse data collection with monitoring. Program managers need to exercise the discipline to routinely collect, analyze, and reflect on information about their program at both the activity and outcome levels. Continuous, disciplined analysis of key program dynamics can result in profound improvements in relevancy, effectiveness, sustainability, and impact.

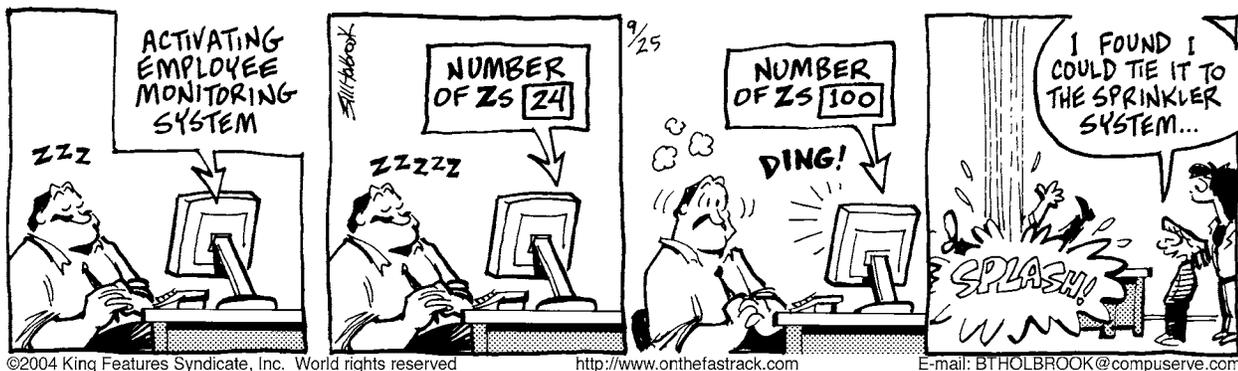
In development programs, where the needs are more static, monitoring often addresses the question, “Are we doing what we said we would do?”

Both monitoring and evaluation use data to inform decision-making and contribute to improved strategies.

Given the dynamic context of many conflicts, a more pressing question in peacebuilding is, “Are we doing what needs to be done?” In identifying what data to collect, consider which decisions the data will inform. Are there additional decisions to be made that lack data?

What is context monitoring?

Peacebuilding often takes place in a very fluid environment where circumstances can improve or deteriorate quickly. Monitoring the context helps peacebuilding practitioners anticipate changes, make proactive programmatic shifts, and ensure the safety of participants, partners, and staff. Context monitoring is the continuous updating and refinement of the conflict assessment.



How do we monitor the context?

Timely and continuous conflict analysis is both present- and future-focused and is essential in designing strategic interventions. Ideally, conflict assessments are living documents that are continuously updated and modified according to developments and changes. Repeating a complete conflict assessment every month is neither feasible nor desirable. However, there are often flashpoints or triggers that should be followed closely. Factors to consider are those that could escalate the conflict as well as those that could deescalate the conflict. Where the conflict has become cyclical, recognizing patterns can help anticipate periods of increasing violence.

Context monitoring is the continuous updating and refinement of the conflict assessment.

The frequent and repeated need for information makes many of the more formal means of data collection overly cumbersome. Context monitoring relies heavily on key informants - people who have unusual access to information and people who are engaged in analysis. This includes a wide range of people such as civil servants, university professors, editors and journalists, diplomatic officials, advisors attached to international organizations, and local leaders. Ad hoc meetings of people who are trying to make sense of recent developments and/or discrete advisory groups are

also sources of information. In addition, where possible, someone from within the organization should check in periodically and individually with people from the different groups in conflict.

For some organizations, monitoring the context is their only peacebuilding program intervention or strategy. Examples of context monitoring include the OSCE High Commissioner on National Minorities, elections monitoring, cease-fire monitoring,¹¹ documentation of human rights abuses, early warning systems and many others.

A number of organizations specialize in early warning systems and make their information and projections available to others in the field. A comprehensive list and general description of conflict assessment and early warning tools can be found in *Conflict-Sensitive Approaches to Development, Humanitarian Assistance and Peacebuilding* (available online at http://www.conflictsensitivity.org/resource_pack.html). Two early warning organizations merit special mention: Swisspeace and the International Crisis Group.

- Swisspeace provides a periodic update on certain countries using the FAST methodology. The purpose of these updates is to provide development agencies, foreign ministries, international organizations, and NGOs with periodic risk assessments and early warnings. FAST continuously monitors and tracks social, political, and economic developments since these can indicate the potential for instability and violence.
- The International Crisis Group (ICG) produces high-quality thematic assessments that sometimes also serve an early warning function. Whereas FAST attempts to synthesize many events from several different fields, ICG provides a more in-depth assessment of specific dynamics, issues, or conflicts such as land tenure or terrorism.

These sources can be valuable supplements for busy practitioners, but they can never fully substitute for thorough local information. Practitioners must develop and maintain their own sources of information about the changing contexts in which they work.

Peacebuilding programs that neglect to routinely update their assessments do so at their peril. The interval between updates depends largely on the volatility of the conflict situation. Monthly updates are not unusual in a rapidly changing environment. Following an updated assessment of the context, a number of programmatic questions need to be reconsidered, such as:

- Are our assumptions about the context still valid?
- Are our interventions still strategic?
- Are there new opportunities?

¹¹ To learn more about the OSCE High Commissioner, see *Peacebuilding: A Field Guide*, edited by Luc Reychler and Thania Paffenholz (Lynne Rienner Publishers, 2001)

- Are there other activities that previously were not viable that might work now?
- What other approaches should we consider?
- If everything else has changed, why continue to do the same things we've always done?

Safety and security are very real considerations in almost all types of peacebuilding, particularly those that bring people in conflict into shared spaces. Whether working with domestic violence or in war zones, peace workers also need to ensure the safety of participants, partners, and staff. In war zones, the level of security frequently changes and up-to-date information is essential to keeping people out of harm's way. In most situations this means daily updates, although in some cases, hour-to-hour monitoring is needed.

Implementation monitoring tracks how the project is running and provides key information for decisions by project managers and participants as well as information for reports to supporters and stakeholders.

What is implementation monitoring?

Implementation monitoring tracks how the project is running and provides key information for decisions by project managers and participants as well as information for reports to supporters and stakeholders. It contributes to keeping the project moving forward. Most program managers tend to do this by comparing planned activities to implemented activities and the resulting outputs. The following example is typical of many implementation monitoring efforts where the focus and the information are limited to the activities or outputs, rather than higher level changes.

Example of Implementation Monitoring Data

| Objective | Activity | Assumption | Data collected about activities |
|---|-------------------------|--|--|
| Local leaders from all communities collaboratively resolve inter-community disputes | Train community leaders | Women and men who receive training will use their new skills to resolve disputes | Number of men and women trained Number of training workshops held by location |

Program managers could use this information for a number of important considerations.

Connection between Data and Decisions

| Data | Decisions the data informs |
|--|---|
| Number of men and women trained | <ul style="list-style-type: none"> • Are we training enough people? • Are we meeting the project deliverables? • Are enough women being trained? |
| Number of training workshops held by location and date | <ul style="list-style-type: none"> • Were workshops held in the targeted locations? • Were enough workshops held? • Were the workshops held at the right time? |

Monitoring progress toward results implies monitoring progress toward change.

Data could also be collected on whether the individuals trained were part of the old or new leadership in the communities by tracking how long they have held their leadership positions. If age or religion are important factors, these could be tracked to ensure that the appropriate groupings are being included.

Programs are usually accountable to a number of constituents: participants, donors, partners, supporters, and the larger organization. Data from implementation monitoring can help explain to these constituents what has been implemented. Traditionally, reporting to donors focuses on planned versus accomplished outputs. Donors may want to know if the inputs were sufficient in both quantity and quality. Did the inputs result in the anticipated outputs? What activities have taken place? Who participated and how?

How do we monitor progress toward results?

Monitoring progress toward results implies monitoring progress toward change. This goes beyond simply reporting on planned versus actual activities and outputs. Here we want to use available data on the objectives and related indicators to inform decisions. Training in peacebuilding programs is an excellent example of an activity for which follow-up information about the application of new skills and knowledge is frequently missing.

Connection between Data and Decisions

| Objective | Activity | Assumption | Data collected about activities | Data collected about changes |
|---|-------------------------|--|---------------------------------|---|
| Local leaders from all communities collaboratively resolve inter-community disputes | Train community leaders | Women and men who receive training will use their new skills to resolve disputes | Number of men and women trained | New types of disputes where trained men and women are getting involved after the training |

This data set informs a different series of decisions that program managers need to make while the program is still being implemented.

Connection between Data and Decisions

| Data | Decision the data informs |
|--|--|
| Types of disputes where trained men and women are getting involved | <ul style="list-style-type: none"> • Does the training prepare people for the types of disputes they are encountering? • Do we need to focus on referral mechanisms for disputes that surpass competencies or jurisdiction? • Are there preventive measures that might be better suited for the types of disputes that are frequently being addressed by local leaders? • Are there types of disputes reserved specifically for women or for men? If so, what are those types? |

Project indicators are signals that the objectives (changes) sought by a project have been reached, but those objectives will often not be reached until the end of the project. How does one monitor progress on these indicators during the course of the project when the changes won't occur until the project ends? Two common ways to cope with this challenge are outlined below.

- Ration the magnitude of the change. For example, X will increase by 10% in year one, 25% in year two, and will reach 100% by the end of year three. Some changes begin to manifest themselves early and continue throughout the program. Even though the full change will not be completed until year three, there may be incremental evidence of success during all three years of the program. Another common rationing device is by geographic region.
- Monitor steps within the process. For example, we may want to monitor specific steps in a larger process. In the example above, a process might include training, outreach, links to referral services, convening the parties, provision of dispute resolution service, quality control, and continuing education for practitioners. The assumption is that substantial progress needs to be made in the first year on training, cultivating referral sources, and outreach if the project is to succeed in three years.

What do we do with testimonials, anecdotes, and personal narratives?

It is important to record, preserve, and appropriately use testimonials, anecdotes, and narratives of personal stories and experiences. These can be wonderful sources of inspiration and insight as well as important opportunities for learning. Moving personal accounts help to put a human face on quantitative data, anchor program outputs in real life situations, and make reports relevant and more interesting. Occasionally, they

enlighten program staff about important new phenomena to monitor. The challenge in using testimonials for monitoring is in having enough testimonials about the same issues and enough additional types of information to validate conclusions and to accurately inform decisions. (For additional information on the liabilities of testimonials, see “Lovers, Haters, and Everybody Else” on page 221 of the Methods chapter.)

Why monitor our assumptions?

Periodically, we need to ensure that the assumptions inherent within the program logic remain valid. This review may occur in the context monitoring where assumptions are based on the situation. At times, it involves routinely collecting and analyzing additional data beyond what is anticipated within the indicators and objectives.

Consider a program designed to reduce the incidence of farmer/herder disputes that escalate into violence in the Sahel. The initial assessment found nomadic pastoralists were frequently in conflict with settled agriculturalists during specific periods of migration. The program diligently tracked the numbers of people trained and serving as third-party neutrals, the use of third-party neutrals, efforts at publicizing prevention measures, the types of disputes, and the outcomes resulting from farmer/herder disputes.

Had the program staff members also monitored changes in the practices of the target groups, they would have found significant changes that would require them to reconsider their programming choices. Instead, they assumed the disputes would continue to be between nomads and fixed residents. A more in-depth look at who was participating revealed that many who were formerly strict agriculturalists were now practicing animal husbandry as well. Many disputes were no longer between itinerant visitors during specific times of the year but between neighbors all year long.

Should information be collected for each group involved?

The effort to collect information as part of a continuous monitoring practice should take careful consideration of the different groups involved. In particular, data needs to be disaggregated by gender and by other dimensions that define the participants in the conflict such as area of origin, age, religion, nationality, identity, and ethnicity. The initial assessment should indicate the areas of concern most relevant to the conflict. See page 216 of the Methods chapter for further information on data disaggregation.

Periodically, we need to ensure that the assumptions inherent within the program logic remain valid.

How does monitoring fit with the logical and results frameworks?

At the design stage, the main components of a detailed monitoring and evaluation plan are summarized in the logical framework under columns labeled objectively verifiable indicators, means of verification, and assumptions. Having a detailed monitoring plan makes completing the logical framework much easier. Results frameworks often have a separate document which details the components of the monitoring and evaluation.

Continuous, timely information from the program about the changing context can facilitate rapid donor approval of needed modifications.

What do we do when monitoring indicates that we need to make a major programmatic shift?

In a very dynamic conflict, the timing of donor and administrative cycles rarely correspond with needed programming developments and modifications. The annual or three-year strategic plans that are common in development work risk becoming out of date in a dynamic conflict before they are even finalized. Proposals submitted to donors become irrelevant quickly when circumstances on the ground change dramatically.

The fluid context in a conflict may require an additional type of reporting. Many donors monitor the context independently and appreciate knowing how their partners view changes in the context. Continuous, timely information from the program about the changing context can facilitate rapid donor approval of needed modifications. Donors are much more interested in achieving higher level objectives than they are in implementing certain activities, particularly those activities that are no longer relevant. Candid and open discussions are the best means for developing acceptable alternatives within existing partnerships.

Example: Restorative Justice for Youth Program Monitoring Plan¹²

Consider a family group conferencing (FGC) program in New Zealand. Family group conferencing is part of a juvenile restorative justice program. The process was based on indigenous community justice practices of the Maori.

Goal: To heal the damage to victims, offenders, and communities caused by youthful offending.

Objective: Involve those most affected by the youth offenses to determine appropriate responses.

¹² Morris, Allison and Gabrielle Maxwell. 1998. "Restorative Justice in New Zealand: Family Group Conferences as a Case Study." *Western Criminology Review* 1 (1). [Online]. Available: <http://wcr.sonoma.edu/v1n1/morris.html>.

Monitoring plan a

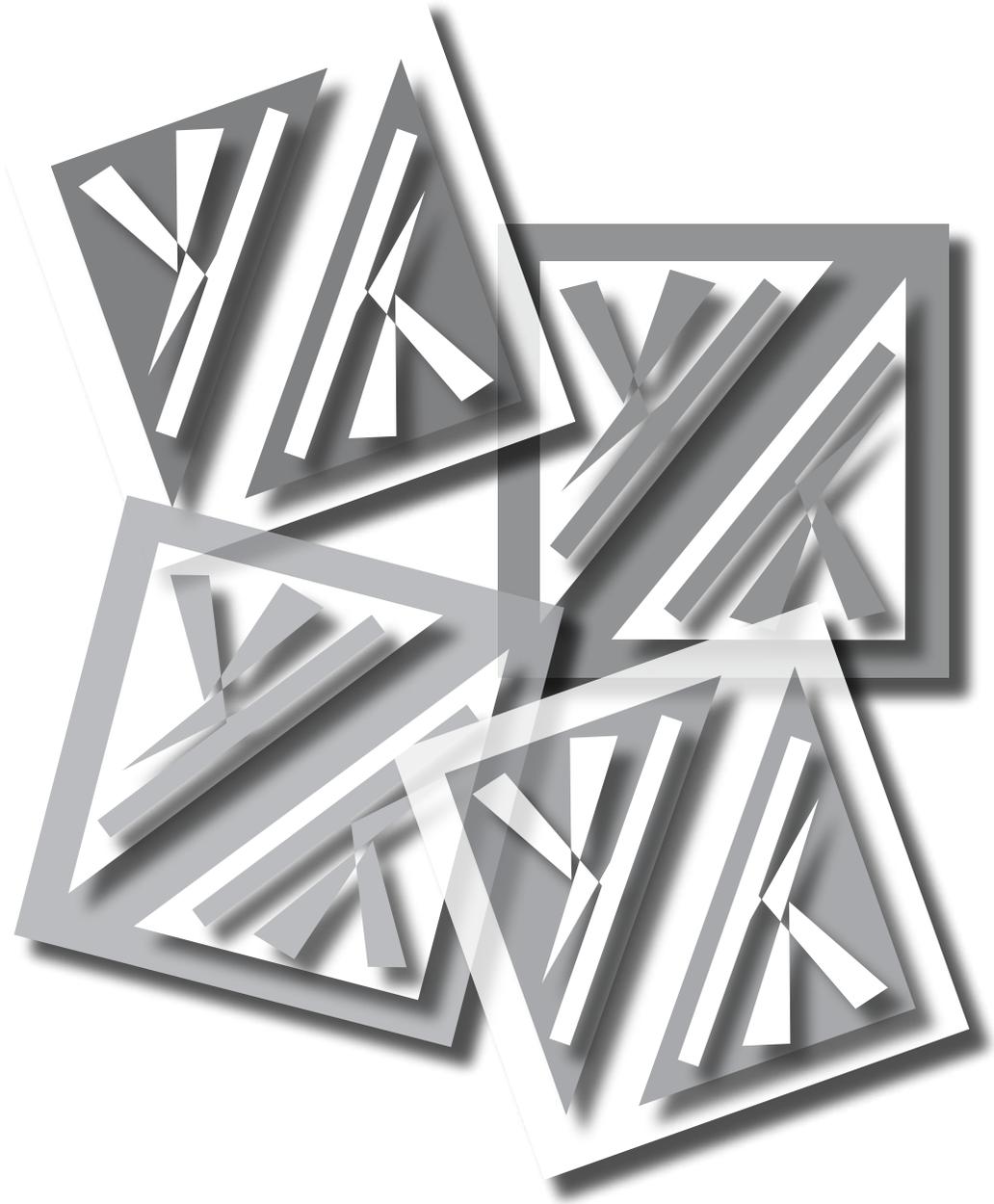
| Outcome-related data | Decisions that data informs | Frequency of data collection | Who collects data and how? |
|--|---|--|--|
| % of male/female offenders who feel involved in the restorative process | Does the process open opportunities for youth involvement? (Program management team) | Collected at the end of each conference | Self-reporting, written questionnaire |
| % of male and female victims who three months later still feel their interests have been addressed | Is there a need for follow-up? (Individual program managers) Are there ways to make the agreements more durable? (Program management team) | Collected at the end of each conference and three months later | Self-reporting written questionnaire, and follow-up phone call later |
| Number/Percentage of potential cases that chose FGC | Are more or less people inclined to choose FGC? (Program outreach team) | Quarterly | Service Coordinator from database |

Activity: Conduct a family group conference.

Monitoring plan b

| Output-related data | Decisions that data informs | Frequency of data collection | Who collects data and how? |
|--|---|-------------------------------------|-----------------------------------|
| Number of families offered FGC | Are outreach efforts contacting enough people? | Quarterly | Service Coordinator from database |
| Number of families offered FGC that complete the process | What percentage of the families that start the process finish it? | Annually | Service Coordinator from database |

In addition to outcomes, this program may also need to monitor the assumptions implicit in its theory of change, which in this example is a combination of the individual change model and the healthy relationships model. One assumption is that restorative approaches reduce recidivism – in other words, there is a belief that young offenders who go through the program will be less likely to offend again. Another assumption is that strong social connections will prevent youth from offending people they know. Recidivism, even though not explicitly stated within an objective, is an important behavior to monitor in order to ensure that program assumptions are still valid. A large increase in recidivism might require the program to rethink its objectives and activities.



EVALUATION
INTRODUCTION

INTRODUCTION

“All truths are easy to understand once they are discovered; the point is to discover them.”

- GALILEO GALILEI

Evaluation is the systematic acquisition and assessment of information gathered on specific questions to provide useful feedback for a program, organization, or individual.

Evaluation preparation, stage one, reviews the decisions that need to be made regarding the evaluation during the project design.

Evaluation is the systematic acquisition and assessment of information gathered on specific questions to provide useful feedback for a program, organization, or individual. It is “a methodological area that is closely related to, but distinguishable from more traditional social research. It utilizes many of the same methodologies used in traditional social research, but because evaluation takes place within a political and organizational context, it requires group skills, management ability, political dexterity, sensitivity to multiple stakeholders and other skills that social research in general does not rely on as much.”

Evaluation is commonly thought to serve two purposes: learning and accountability. The two purposes are not separate; in fact, they overlap and reinforce each other significantly, since to be accountable implies the requirement to learn from success and failure.

Monitoring and evaluation are often confused with each other. They are related but distinct activities with a common overarching goal: to provide information that improves decisionmaking and facilitates learning. Evaluation generally offers a more in-depth analysis because it asks why something happened, whereas monitoring is more often simply providing basic data. More information on the distinctions between monitoring and evaluation can be found in the Monitoring chapter, page 81.

The evaluation portion of this manual is laid out in a three-stage process.

Stage 1: Evaluation Preparation

Stage 2: Evaluation Management

Stage 3: Evaluation Utilization

Evaluation preparation, stage one, reviews the decisions that need to be made regarding the evaluation during the project design. Knowing such things as why you are doing the evaluation, the evaluation audience, evaluation type, the evaluator’s role and qualifications, timing, and budget will enable the project team to plan wisely and maximize the benefit from the evaluation.

The decisions made in the evaluation preparation stage lay the groundwork for evaluation management, stage two. In stage two, the terms of reference and evaluation plan are described. These topics, as well as frequently asked questions on working with external evaluators and a discussion on strategies for common evaluation pitfalls, are all addressed in the Evaluation Management chapter.

Stage three discusses utilizing the evaluation at all levels of the project, the organization, and the wider conflict transformation field. While evaluation utilization may be labeled as the third stage, this does not truly reflect when the thinking on utilization actually begins. For example, how the findings and recommendations of the evaluation will be used is tied closely to why the evaluation is being conducted, which is a stage one decision. The chapter on evaluation utilizations page 178 covers how to ensure that evaluations foster learning and, consequently, how they generate practical new applications at all levels.

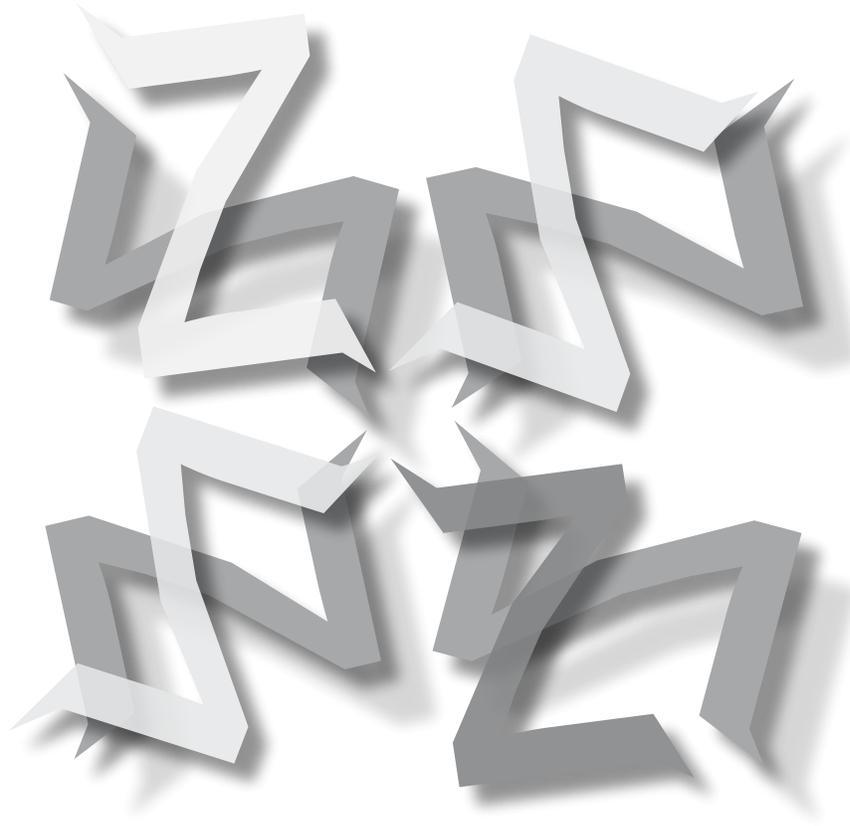
The following table provides a “quick and dirty” summary of key aspects of each of the stages.

In stage two, the terms of reference and evaluation plan are described.

Stage three discusses utilizing the evaluation at all levels of the project, the organization, and the wider conflict transformation field.

Evaluation Stages Overview

| Stage | Topic | Activities | Timing | Deliverables |
|-------|------------------------|---|--|---|
| 1 | Evaluation Preparation | <p>Make core decisions for the evaluation on:</p> <ul style="list-style-type: none"> • evaluation objectives • audience • evaluation type • role • approach • scope • evaluators • timing • budget | During project design | Project design documents including key evaluation points |
| 2 | Evaluation Management | <p>Finalize core decisions</p> <p>Develop terms of reference</p> <p>Recruit evaluation team</p> <p>Create evaluation plan</p> <p>Implement evaluation</p> | Starts 4-6 months before evaluators are expected to collect data | <p>Evaluation section of donor proposal documents competently completed</p> <p>Terms of reference</p> <p>Evaluation deliverables (e.g., report, presentation)</p> |
| 3 | Evaluation Utilization | Reflect, apply, generalize, and share new knowledge identified by the evaluation within the organization and with the broader peacebuilding field | Many utilization decisions will occur in the evaluation preparation stage, while implementation will occur in the latter third of the evaluation cycle | <p>Utilization plan</p> <p>Possible outreach efforts:</p> <p>Academic journals</p> <p>Concise summaries for the public</p> <p>Lessons learned documents</p> |

**STAGE ONE****EVALUATION PREPARATION**

This section contains:

1. The Actors Involved in Evaluation Preparation
2. The Core Preparation Decisions and How They Relate to Each Other
 - *Evaluation Objectives*
 - *Audience*
 - *Type of Evaluation*
 - *Evaluator's Role*
 - *Evaluation Approaches*
 - *Evaluation Scope*
 - *Type of Evaluator*
 - *Timing of the Evaluation*
 - *Budget*
3. Length of the Evaluation Preparation Process

INTRODUCTION

“The first step to getting the things you want out of life is this: Decide what you want.”

- BEN STEIN

Evaluation preparation is part of the conflict transformation project design process. By making the core decisions at this stage, it not only aids the proposal process by providing clear answers to the evaluation section within donor proposal forms but also enables the project budget to incorporate a reasonably accurate financial projection to fund the evaluation. Making key decisions now also saves time later, when the project is underway, since these decisions inform many of the evaluation terms of reference. Most importantly, it reinforces to the project team the areas where more information would allow for improved decisionmaking and the value of continuous learning.

Preparing an evaluation is much like designing a conflict transformation project. Start with what the project team wants to achieve and for whom, then address the practicalities of how, when, who, and at what cost. As illustrated on page 99, Evaluation Preparation Decision Flowchart, the decisions are interdependent such that changing one affects many of the others.

Evaluation preparation starts with determining what the team wants to learn from an evaluation. The outcome of this deliberation will be a set of evaluation objectives that will achieve the learning goal. This decision about what the team wants to learn is often tied very closely to who the audience is for the evaluation. In some situations, perhaps as a result of donor requirements or organizational needs, an agency may choose to start the evaluation preparation process with determining the audience first and then shifting to the evaluation objectives.

Once it has been determined why the team is doing the evaluation and who is the primary audience, other more practical decisions can be made: evaluation type, evaluator role, methods, type of evaluator, and timing. Finally, it is important to ensure that the budget is adequate for the process that has been designed. Very often, the development of the evaluation budget functions as a reality check requiring the organization to revisit earlier decisions. The two most common parts of the evaluation process to reconsider in terms of their impact on the budget are the evaluation objectives and the methods. Leaving the financial decisions until the core evaluation design decisions have been made enables greater creativity, which can be beneficial since it provides time for deeper reflections on what is truly needed by the program.

Evaluation Preparation is part of the conflict transformation project design process.

The benefits of making these core evaluation decisions during the project design stage are immediate. Most donor proposal formats require an explanation of the intended evaluation. Much of that explanation will come as a result of working through the decisionmaking flowchart. In turn, the project budget will accurately reflect the financial needs of the evaluation.

Who should be involved in the evaluation preparation?

Without question, the project team must lead the decisionmaking in the evaluation preparation. If the evaluation is to contribute to learning and improvement it must be grounded in the realities of the project, the organization, and those intent on learning. Design, Monitoring, and Evaluation (DM&E) Technical Assistants are helpful in providing the pros, cons, and ramifications of different options within each decision. The outcomes of these conversations need to be incorporated into the proposal and budget process; hence, including proposal writers in the process is also helpful.

Another option is to hire an evaluator whose engagement would start at the project design phase. Ideally, this would be the same person who would later return to conduct the evaluation(s). This person's role could include any or all of the following:

- Assisting the project team in making the evaluation preparation decisions.
- Contributing to the project logic development, indicators, and means of verification.
- Establishing the monitoring system and training staff in its implementation.
- Conducting the baseline.
- Conducting a mid-term (formative) evaluation. More information on formative evaluations may be found on page 110 of this chapter.
- Conducting a final (summative) evaluation. More information on summative evaluations may be found on page 110 of this chapter.

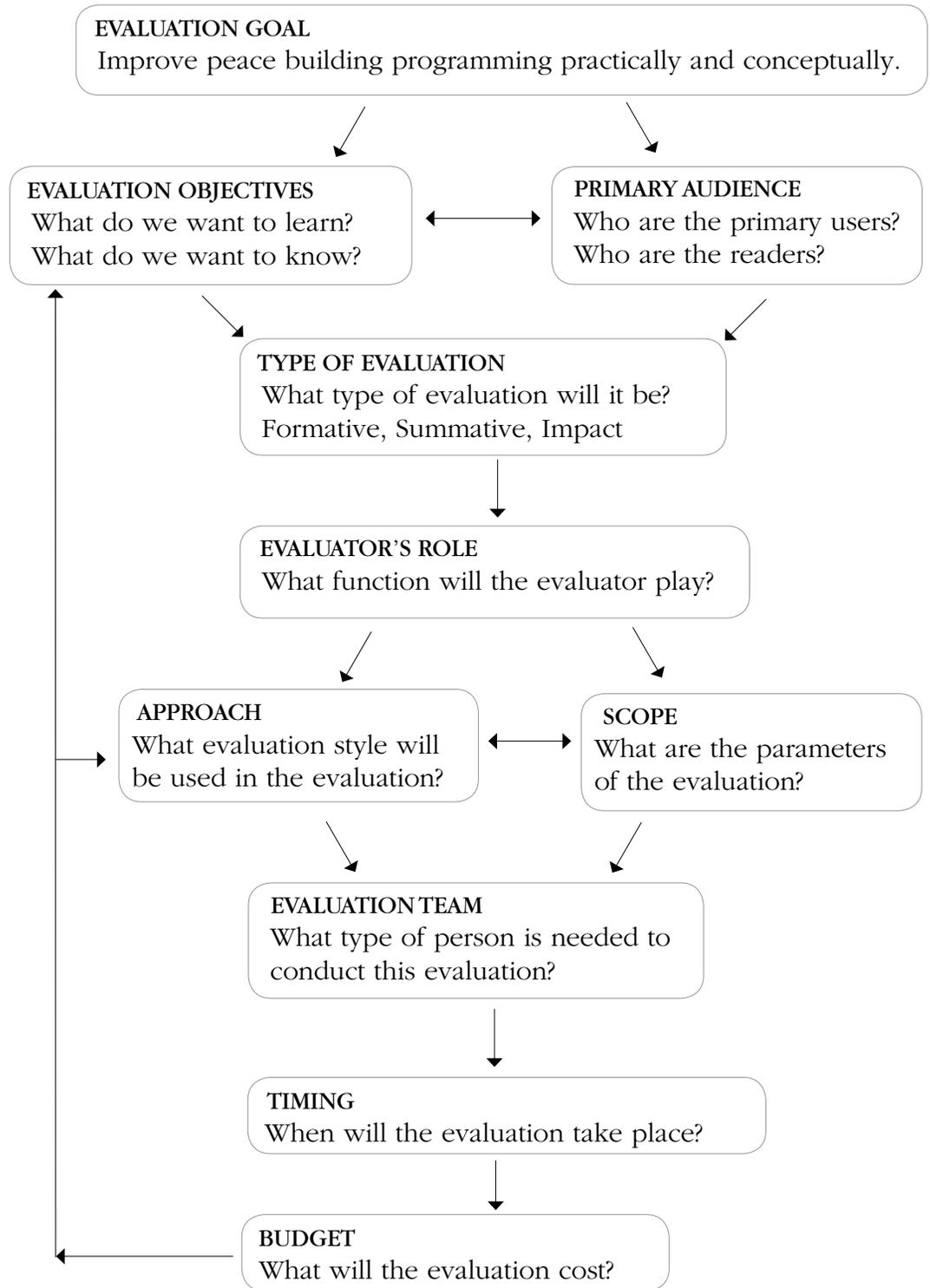
If the evaluation is to contribute to learning and improvement it must be grounded in the realities of the project, the organization and those intent on learning.

Preparation for the evaluation occurs in the project design stage

When does the evaluation preparation occur?

Preparation for the evaluation occurs in the project design stage. The decisions made should be specific enough to provide direction while remaining flexible enough to adapt to changing needs as the project progresses. These decisions will be reviewed and altered, if necessary, when the evaluation is implemented.

Evaluation Preparation Decision Flow Chart



I. DECISION: EVALUATION OBJECTIVES

What do we want to learn?

There are many different areas in which an evaluation can facilitate greater learning. Selecting the areas for learning is part of the process of determining the evaluation objectives. An evaluation objective is the criteria by which a project will be evaluated. These objectives can range from the traditional, such as identifying results, to the less traditional, such as determining if the activities of a project are in alignment with the organizational vision. Spending time identifying what information the project management needs to inform decisions, improve future performance, and understand more deeply how and why the project is effecting change sets the stage for an evaluation to contribute to the organization and its mission.

An evaluation objective is the criteria by which a project will be evaluated.

In some cases, there may be donor requirements that need to be included in the evaluation objectives. Frequently, donor requirements can be modified or expanded to include other issues by discussing options with the donor.

What are the existing criteria or frameworks we can use to guide our thinking?

The Conflict Transformation Evaluation Framework¹³ used in this manual compiles the possible evaluation objectives and is grouped into three main themes. The first theme explores why and how the agency is conducting this type of intervention. The second theme considers how well the intervention was implemented, and the third covers the results and how long they will last.

Conflict Transformation Evaluation Framework

| Theme | Evaluation Objective |
|--|--|
| Why and how is the agency conducting this intervention? | Appropriateness Consideration Strategic Alignment |
| How well was the intervention implemented? | Management and Administration Cost Accountability Implementation Process Appraisal |
| What were the results of the intervention and how long will they last? | Output Identification Outcome Identification Impact Assessment Adaptability of Change |

¹³ Modified from *The Evaluation of Conflict Resolution Interventions: Framing the State of Play* by Cheyanne Church and Julie Shouldice, (INCORE, 2002) [hereinafter “Church and Shouldice, *Framing the State of Play*.”]

An evaluation does not need to cover all evaluation objectives.

An evaluation does not need to cover all evaluation objectives. Rather, the framework's "purpose is to offer stakeholders an instrument ...to organize their thinking and constitute the aims, objectives and terms of" an evaluation.¹⁴ Practitioners should also not feel limited by the framework, and if there are other objectives that might benefit learning, those should be utilized.

An overview of two related concepts has been provided because there is a great likelihood that practitioners will come across them in their evaluation efforts. The first on page 105 is the OECD Development and Cooperation (DAC) criteria, which are widely used in the humanitarian field. Although called criteria, they can be used in the same way as evaluation objectives. How the five criteria relate to the Conflict Transformation Evaluation Framework is also described in that section.

The second concept, developed by the Reflecting on Peace Practice Project (RPP), found on page 107, is specifically focused on evaluating a project's achievement towards peace writ large. Although also called criteria, these are of a slightly different nature than the DAC criteria because they cannot replace evaluation objectives. These criteria could be used as part of an evaluation whose objective is impact assessment if the definition of impact related to peace writ large. At the time of writing, the criteria are still in the testing stage, and one should check with RPP to get the most up-to-date rendition.

Can the evaluation objectives in the conflict transformation framework be explained further?

APPROPRIATENESS CONSIDERATION reviews whether the intervention strategy is the best for the situation and desired goal. Included in this objective is a review of the theory of change and whether the strategy is based on the needs or opportunities of the target population as determined by the conflict analysis.

This objective is useful for those operating in a rapidly changing environment who want external input into the relevance of their strategy. It also provides commentary on the quality of the conflict analysis and can be helpful in facilitating additional thinking on theories of change. Appropriateness consideration is most usefully included in a formative evaluation since readjustments are immediately possible, although there may be scenarios where it is valued in summative as well. To fully understand the effectiveness of a strategy, this objective should be paired, at the very least, with Outcome Identification.

STRATEGIC ALIGNMENT addresses whether the activities of an organization are in line with the organization's mission and principles.

¹⁴ Church and Shouldice, *Framing the State of Play*.

For an organization to effect the change articulated by its mission, it must allocate all its resources and attention to that change rather than scattering its efforts among many different valuable but unrelated projects.

This objective helps project teams ensure that their work embraces the core principles of the implementing organization. If gender equality is considered a core principle of an organization, the evaluation would examine if that principle was being incorporated into the project. It also considers if projects are within the mandate of an organization. For instance, should an organization with a mission to reduce violent conflict start implementing projects on HIV/AIDS or girls education? This objective can also look at the country office level to see if the projects within one office are strategically aligned.

MANAGEMENT AND ADMINISTRATION considers how well the project was organized and run. It covers the supervision, communication, and implementation aspects of the program.

This objective helps provide input into better management systems to more effectively implement projects. It can be very useful as part of a formative evaluation for project teams that have not worked together previously or in offices that have recently been opened. It can also provide some input into why a project is not advancing the way it was expected.

COST ACCOUNTABILITY reviews the manner in which funds were utilized and accounted for by the organization. Depending on the conflict context, there will not always be an expectation of achieving the desired ends for the lowest cost.

Often reviewed along with the management and administration objective, cost accountability provides insight into financial decisions and processes. This objective looks at who has responsibility for financial decisions and how they affect the implementation of the project. It also considers whether other choices could be made that would avoid harm to the project yet save resources. Although there are overlapping elements between these two objectives, cost accountability should not be confused with a financial audit because it does not review technical accounting practices. This objective is useful for projects that are either extremely multi-faceted or highly dynamic, and require many project changes.

IMPLEMENTATION PROCESS APPRAISAL examines the quality of the conflict transformation techniques used in the implementation of the project. This objective considers the merits of the “process” utilized in the intervention. The process involved in doing a training, for instance, includes the agenda development, participant selection, and the training techniques used such as creating a safe space for learning, utilizing games to highlight key points, or providing information in different formats.

This is a valuable evaluation objective for a pilot project that is using an innovative model. It provides insights into the mechanics of what is working and what is not, and helps foster the ongoing development of the model. The benefits of this objective can be increased significantly if it is included in a formative evaluation. This objective is also useful if a model is being adapted to a new cultural context or the project staff members are still developing their skills.

OUTPUT IDENTIFICATION considers the immediate, often-tangible results of the activities undertaken. It tallies the number of “things” that have been produced by the project’s work thus far. These things might include the number of people trained, the number of pamphlets printed, the number of mediation cases handled, or the number of radio minutes produced.

Donors, as part of their accountability requirements, often mandate this evaluation objective. It would rarely be the sole evaluation objective, however, since the information it provides would normally be insufficient to inform and shape decisionmaking. The exception would be a mid-term (formative) evaluation done so early in the project that outputs are all that can be realistically expected. Coupled with outcome identification (below), this objective can provide a great deal of useful information.

OUTCOME IDENTIFICATION explores the changes that result from the project’s activities. These changes include shifts in processes as well as the unintended positive and negative effects. Examples of outcomes range from a decrease in violence to an increase in collaboration between communities.

This evaluation objective is highly informative for project teams both during and after a project. It provides data on the success of the project to date in terms of the changes achieved. If utilized in a formative evaluation, the timing of that evaluation should ensure that it is reasonable to expect outcome-level changes. It is commonly coupled with output identification.

IMPACT ASSESSMENT investigates the consequences or changes resulting from an intervention in the conflict situation or in a component of that situation. The scope of the impact assessment varies with the scale of the project – from “peace writ large” down to a local community – and should include the unintended positive and negative effects. This assessment often involves identifying the transfer¹⁵ of changes from the target group to others.

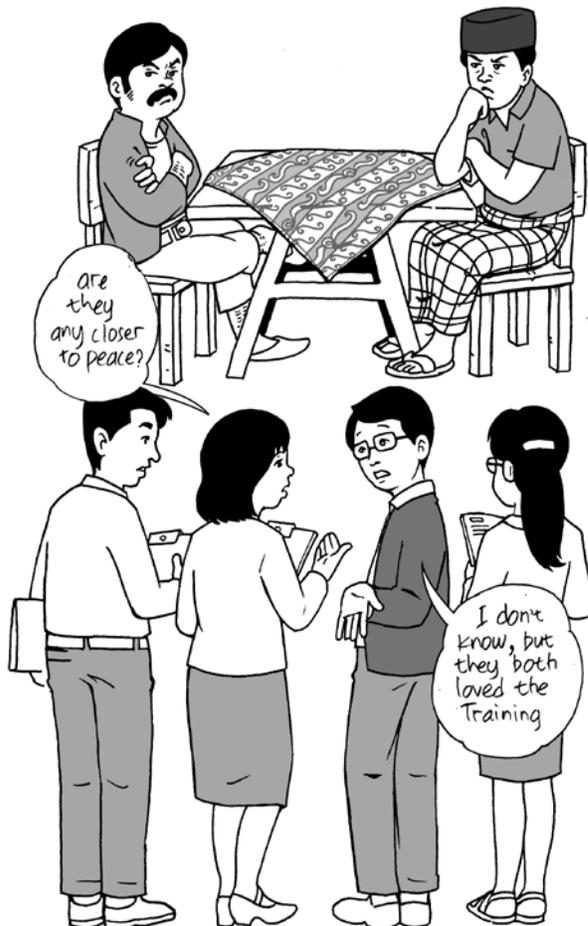
Impact assessments are difficult to conduct and often require greater time and financial backing than other evaluation objectives. In the field of conflict transformation at the time of writing, few methodologies have

¹⁵ For further information on the idea of transfer, see *The Evaluation of Conflict Resolution Interventions, Part II: Emerging Practices & Theory* by Cheyanne Church and Julie Shouldice, (INCORE, 2003) [hereinafter “Church and Shouldice, *Part II.*”]

been developed to adequately answer the impact challenge. Adopting this evaluation objective will require that resources be allocated to allow for both new methodology development and the evaluation implementation. If successful, an impact assessment objective would be highly informative to the project team. This objective applies to summative and impact evaluations.

ADAPTABILITY OF CHANGE reviews whether the changes (outcomes or impact) created by the project can adapt over time to shifts in the context and to different stresses and demands. This objective implies more than sustainability of results throughout a phase of the conflict. Instead, it suggests that the results or changes should evolve appropriately to meet the demands of a new phase in the conflict.

This evaluation objective would have significant value for program design; however, as with impact assessment, it is a new area of evaluation whose methodologies have yet to be tested. If adaptability was to be the focus of an evaluation, sufficient resources to develop the methodology would need to be allocated. A much longer time period than the average project cycle would also need to be used.



Illustrated by Ariv Russanto, Windi, Wahyu S., Ary WS.
Creative team Search for Common Ground in Indonesia

Do we need to use all of the evaluation objectives in the framework?

It is not necessary to include all of the evaluation objectives in every evaluation. Project teams should select the ones that will help them improve the project. How many evaluation objectives are feasible is dependent on the size and scope of the program, the difficulty of data collection for each objective, the size of the budget, and the type and number of lines of inquiry included within each objective. Additional information on the lines of inquiry is outlined in Chapter 9 Evaluation Management page 142.

If the information cannot be utilized in any way, it is merely interesting, which indicates that this evaluation objective should be moved to the bottom of the priority list

Curious practitioners will often find all of the potential evaluation objectives interesting. At the end of a discussion about what the team wants to learn, the complete list of objectives may appear; however, evaluating all of them is often well beyond the financial means of the evaluation. One way to help decipher which objectives to keep is to ask, “Will the information resulting from inquiry into this evaluation objective be useful or just interesting?” If the information cannot be utilized in any way, it is merely interesting, which indicates that this evaluation objective should be moved to the bottom of the priority list. If the resulting information could effect changes in a project, then it is useful, and the evaluation objective should rise to the top of the priority list.

Consider, for instance, a project with a goal of decreasing violence against Roma in Ireland. One objective is to decrease negative stereotypes of Roma held by young Irish men. The project team has listed process implementation appraisal, strategic alignment, and outcome identification as its evaluation objectives, yet there are insufficient funds to thoroughly explore all three. Asking themselves which of these evaluation objectives would produce interesting or useful information can help to set priorities. This project will continue even if the results of the strategic alignment indicate it does not fall within the mandate of the organization. Understanding that nothing will change as a result of their review helps the project team remove strategic alignment from the list of evaluation objectives. The other two objectives will provide useful information and will be kept.

Advanced Concept

A Comparison: DAC Criteria & Conflict Transformation Framework

The Development and Cooperation Directorate of the Organization for Economic Cooperation and Development (OECD DAC) created the most widely used set of evaluation criteria in the humanitarian and development field. These criteria are utilized in the same way as the evaluation objectives in the Conflict Transformation Framework.

The table below includes the official definitions found in the OECD DAC Criteria column. Note that these are not the definitions of the evaluation objectives within the Conflict Transformation Framework. Since these criteria are sometimes referred to outside the humanitarian/development realm, the table below illustrates how the DAC Criteria relate to the Conflict Transformation Framework. This does not mean that the terms refer to the exact same concepts; instead, it shows where they are conceptually similar. In other words, “efficiency” does not translate exactly to “management and administration”; rather, those concepts are broadly aligned.

Strategic alignment and implementation process appraisal from the Conflict Transformation Framework are not included in the chart because there are no equivalent concepts in the DAC Criteria.

Distinguishing OECD DAC Criteria and Conflict Transformation Framework

| OECD DAC¹⁶ Criteria | Conflict Transformation Framework |
|--|--|
| <i>RELEVANCE. The extent to which the objectives of a development intervention are consistent with beneficiaries’ requirements, country needs, global priorities and partners’ and donors’ policies.</i> | <i>Appropriateness Consideration</i> |
| <i>EFFECTIVENESS. The extent to which the development intervention’s objectives were achieved, or are expected to be achieved, taking into account their relative importance.</i> | <i>Output Identification Outcome Identification</i> |
| <i>EFFICIENCY. A measure of how economically resource/inputs (funds, expertise, time, etc.) are converted to results.</i> | <i>Cost Accountability Management & Administration</i> |
| <i>IMPACTS. Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.</i> | <i>Impact Assessment</i> |
| <i>SUSTAINABILITY. The continuation of benefits from a development intervention after major development assistance has been completed. The probability of continued long-term benefits. The resilience to risk of the net benefit flows over time.</i> | <i>Adaptability of Change</i> |

¹⁶ Development Assistance Committee (DAC) Working Party on Aid Evaluation, *Glossary of Key Terms in Evaluation and Results Based Management*, Organization for Economic Co-operation and Development (OECD), 2002, <http://www.oecd.org/dataoecd/29/21/2754804.pdf>.

Five Criteria of Effectiveness for *Peace Writ Large*¹⁷

The Reflecting on Peace Practice (RPP) process produced five criteria of effectiveness (listed below) by which to assess, across a broad range of contexts and programming approaches, whether a program is (or is not) having meaningful impact at the level of “peace writ large.” These criteria can be used in program planning to ensure that specific program goals are linked to the large and long-term goal of peace writ large. They can be used during program implementation to reflect on effectiveness and guide mid-course changes, and as a basis for evaluation after the program has been completed.

It is important to note that the criteria are in the process of being tested in terms of gathering data to confirm that they provide the right information for decisionmaking with regard to peace writ large.

1. ***The effort contributes to stopping a key driving factor of the war or conflict.*** *The program addresses people, issues, and dynamics that are key contributors to ongoing conflict.*
2. ***The effort contributes to a momentum for peace by causing participants and communities to develop their own peace initiatives in relation to critical elements of context analysis: what needs to be stopped, reinforcement of areas where people continue to interact in non-war ways, and regional and international dimensions of the conflict.*** *This criterion underlines the importance of “ownership” and sustainability of action and efforts to bring about peace, as well as creating momentum for peace, involving more people.*
3. ***The effort results in the creation or reform of political institutions to handle grievances in situations where such grievances do, genuinely, drive the conflict.*** *Peace practice is effective if it develops or supports institutions or mechanisms to address the specific inequalities, injustices and other grievances that cause and fuel a conflict. This criterion underlines the importance of moving beyond impacts at the individual or personal (attitudinal, material or emotional) level to the socio-political level. This criterion must be applied in conjunction with a context analysis identifying what the conflict is NOT about and what needs to be stopped. To reform or build institutions that are unrelated to the actual drivers of a specific conflict would be ineffective.*

¹⁷ Taken from Reflecting on Peace Practice: Five Criteria of Effectiveness Working Paper, <http://www.cdainc.com>.

4. ***The effort prompts people increasingly to resist violence and provocations to violence.*** *One way of addressing and including key people who promote and continue tensions (e.g., warlords, spoilers) is to help more people develop the ability to resist the manipulation and provocations of these negative key people.*
5. ***The effort results in an increase in people's security and in their sense of security.*** *This criterion reflects positive changes both at the socio-political level (in people's public lives) and at the individual/personal level, as people gain a sense of security.*

These criteria can best be thought of as intermediate-level benchmarks of success applicable to the broad range of peace work being done.

THE CRITERIA ARE ADDITIVE: *The experience gathered through RPP suggests that the effectiveness criteria are additive. Peace efforts that meet more of them are more effective than those that accomplish fewer of the changes.*

II. DECISION: AUDIENCE

Who is the primary audience for the evaluation?

There are two audiences for every evaluation: users and readers. Users are those who will apply the findings and recommendations while readers are those who would be interested in the report to stay informed. Every evaluation must have a user, whereas readers are optional. Identifying the primary audience – the user – generally occurs in tandem with determining the evaluation objectives. One decision is dependent upon the other and, as such, it becomes a blended conversation.

It is common in these discussions to confuse the two groups since there may be as many as five or six different potential readers who would be interested in the evaluation findings, but generally no more than two users of an evaluation. Having more than two users is not impossible, but it may pull the evaluation team in too many directions to be feasible.

When preparing an evaluation, it is important to make a clear distinction between the two groups (users and readers). With regard to users, maintaining this distinction provides direction to the evaluation team. For

instance, knowing the user group helps the evaluators tailor the recommendations, highlight issues of key importance to that audience or select the language in which the final report should be written, such as Bahasa Indonesia or Swahili.

When a donor initiates an evaluation, specific information is often needed to inform its decisions. In such a scenario, the donor is the primary audience. It is possible, however, that the donor will be open to jointly developing the evaluation objectives with the implementing partner such that both would be the audience. Remember, the audience sets the objectives that dictate for whom the evaluation will be useful. Therefore, NGOs should not be surprised when they find evaluations initiated by donors of their work to be interesting but not useful to them. In such a case, the NGO is the reader rather than the user.

Example

IDP-Host Community Conflict

Part 1: An agency recently started working in Sudan, where it is developing a two-year project with a goal of transforming the daily conflicts between leaders among internally displaced persons (IDPs) and host communities in one state (wilayat) from violent means of resolution to cooperative means. The degree of ongoing violence and the rapidly evolving conflict dictate that the project needs to catalyze these changes as soon as possible to cooperative means.

The organization is committed to conducting evaluations as a way to foster learning. As a result, the project team and the wider organization are the primary audience for those evaluations. The donor is very interested in seeing the results of the evaluations but is not the intended user. It is therefore one of the readers. This means that the evaluation objectives will be based on what the users – the project team – need to learn.

The organization recognizes that it has adapted its techniques to a new target group (IDPs) within a new cultural context for the organization. Therefore, obtaining more information on the quality of the process within specific activities would be valuable in making improvements. Since the team wants to learn how well the adaptations to the model have worked and how to improve them, the first evaluation objective is implementation process appraisal.

Because the effectiveness of a model is limited not only to the quality of implementation but also to the degree of progress that it catalyses, the organization has identified output and outcome identification as two additional evaluation objectives.

III. DECISION: FORMATIVE, SUMMATIVE OR IMPACT

What type of evaluation will it be?

There are three different types of evaluation: formative, summative, and impact. Formative evaluations are generally undertaken to determine progress to date and how to improve the project. Summative evaluations provide an overarching assessment of a project’s “value.”

Formative evaluations occur around the middle of a project, and summative evaluations take place near or at the end of a project. Impact evaluations, on the other hand, occur some time after the project is complete to assess its results and, if feasible, the adaptability to change of those results. This manual focuses on formative and summative evaluations.

Deciding Between Formative and Summative

| Benefits of Formative Evaluation | Benefits of Summative Evaluation |
|---|--|
| <ul style="list-style-type: none"> • Contributes to reporting and accountability requirements during the project | <ul style="list-style-type: none"> • Helps project sum up what it has achieved |
| <ul style="list-style-type: none"> • Provides information to improve project before it is too late to make changes | <ul style="list-style-type: none"> • Checks achievements against plans and obligations to donors and participants |
| <ul style="list-style-type: none"> • Provides structured opportunity for reflection so that staff and resources are focused on the project | <ul style="list-style-type: none"> • Provides information as to why and how change occurred |
| <ul style="list-style-type: none"> • Helps clarify program strengths and weaknesses | <ul style="list-style-type: none"> • Generates important information to drive learning |
| <ul style="list-style-type: none"> • Can provide information that assists realignment of project to the changing conflict context | <ul style="list-style-type: none"> • Can create documentation that captures approaches and lessons to be used in the wider organization |

Neither type of evaluation – formative or summative – should be seen as exclusively backward looking.

Example

IDP-Host Community Conflict

Part 2: The project team has determined its evaluation objectives to be 1) output identification, 2) outcome identification, and 3) implementation process appraisal of specific activities. Because the project is a new application of old techniques, it is important for the team to obtain information on its evaluation objectives while there is still time to make adjustments. To do this, the team will conduct a formative evaluation.

Neither type of evaluation – formative or summative – should be seen as exclusively backward looking. In some cases, these exercises can also be used to gather information as part of an ongoing context monitoring effort to inform the project looking forward. Formative evaluations are generally better suited to adding questions that identify needs and opportunities for the future.

Example

IDP-Host Community Conflict

Part 3: The majority of the formative evaluation will focus on the three evaluation objectives identified by the organization. However, in such a rapidly changing context, there is value in allocating a small portion of the evaluation to taking a look ahead to identify the needs and opportunities within the host and IDP communities. This information, added to the other context monitoring data, can inform the organization of changes in the conflict that might be addressed within the existing project or that require a different initiative.

While it is important to understand the differences between the evaluation types, they should not be seen as separate events. Rather, the design, baseline, monitoring, and formative/summative evaluation should all be seen as part of the multi-step evaluation process. The baseline is essential for the evaluation, monitoring data can inform either type of evaluation, and a summative evaluation may contain formative elements if there will be a subsequent rendition of a project.

Finally, there are impact evaluations, which seek to determine the change in the conflict catalyzed by a project. These evaluations almost exclusively look at impact identification and adaptability of change, although other evaluation objectives may also be included. They are implemented at points ranging from several months to several years after the project is finished.

IV. DECISION: EVALUATOR'S ROLE

What role will the evaluators play?

There are three roles that an evaluator can adopt for an evaluation: operative, consultant, and learning facilitator.¹⁸ “It is important for the stakeholders not only to be aware that evaluators’ [roles] can differ but also to take the time to determine which role is best suited for each evaluation.”¹⁹

The operative role involves a more traditional approach to evaluation such that the evaluator remains within the boundaries of implementing the evaluation. This role ends with the writing of the report. An organization that has either a clear system for utilizing evaluations or a summative evaluation based upon quality monitoring information and previous formative evaluations may find an operative role to be effective.

The operative role involves a more traditional approach to evaluation

The role of consultant has a broader remit and encompasses not only what the operative evaluator does but also a significant contribution to the “use” of the evaluation. This contribution may be in the form of developing practical recommendations, facilitating workshops with staff or working with the project team to develop an implementation plan based on the evaluation. This is the most common role for evaluators at present.

The role of consultant encompasses a significant contribution to the “use” of the evaluation.

Finally, the learning facilitator has the broadest mandate of the three. In this role, the evaluator does everything that the other two do while also seeking to link the project learning into both the broader organization and into the development of the next phase of the project. This could include:

- Development of lessons or questions that are applicable beyond the project
- Presentation of results to headquarters
- Development of an ongoing learning system for the project team
- Assistance with new program development

The learning facilitator has the broadest mandate of the three.

¹⁸ Adapted from Church and Shouldice, *Part II*. pp.10-11.

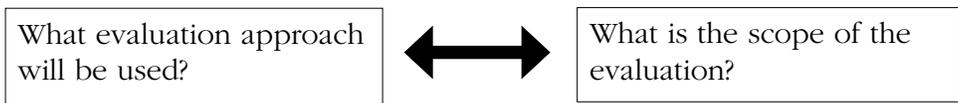
¹⁹ Church and Shouldice, *Part II*.

Example

IDP-Host community conflict

Part 4: After considering the evaluation objectives, the intensive nature of the work plan, and the capacity needs of the staff, the project team chose the learning facilitator role for their evaluator. They want to be certain that the learning from the evaluation informs the project's second half as well as the wider organization. With staff schedules so tight, however, and no in-house experience in utilizing evaluation results, they thought it best to include the learning responsibility in the evaluator's role.

Introduction to Decisions V and VI: Approach & Scope



During the evaluation preparation, there are two key decisions that provide the basis for selecting the evaluation methods. One, the evaluation approach, asks the project team, “What evaluation style will be used?” The other, the evaluation scope, considers the question, “What are the parameters of the evaluation?”

As with all the other decisions at this stage, these answers need to be as specific as possible, while keeping them sufficiently flexible to meet changing needs, particularly as the program draws closer to the evaluation date. Whether the evaluation approach or the scope should be addressed first will be unique to each evaluation and largely dependent upon the evaluation objectives.

Once the approach and scope are determined, the decision will be made regarding which methods to use. Although most methods apply to most approaches, some are less well-suited than others. For instance, if self-evaluation is selected as the evaluation approach, large-scale surveys would not be the most appropriate method to use because they require specific expertise for their development and analysis. Alternatively, if the scope of the evaluation is national-level, generalized conclusions, choosing participatory learning actions and techniques as the methods would not be appropriate in most cases. More information on methods can be found in Chapter 11 Methods page 201.

V. DECISION: EVALUATION APPROACH

What is an evaluation approach?

An evaluation approach provides the framework, philosophy, or style of an evaluation. Consider someone who dresses as a “hippie.” This person does not wear the same clothes everyday, but the style of a hippie informs her/his choices such as where to go shopping, what types of clothes are attractive, or what jewelry to wear with an outfit. An evaluation approach is similar; it is the style of the evaluation.

The Evaluation Approaches section does not include a comprehensive listing of all possible approaches. Rather, it offers a broad range of options that might be useful to conflict transformation practitioners. All but one of the approaches listed have been developed by evaluators and are not specific to the conflict transformation field.

The approaches are not entirely distinct or unique from each other since many have similar concepts. Empowerment evaluation and self-evaluation, for instance, can be considered from the same family. Listings for further reading on the approaches listed may be found at the end of this chapter.

Action Evaluation

Action evaluation is an iterative goal-setting process facilitated by an evaluator throughout the life of a project. Essential to this approach is the assumption that goal setting is a process that continues until the end of the project. With the guidance of an evaluator, project teams set project goals and explore the underlying assumptions and value-basis of why those goals are important. This exploration enables all members of a project team to be fully in step with the purposes of the project and to avoid the pursuit of competing understandings of the goals. Action plans are then created based upon those goals and are used to inform the next goal-setting discussion. In this process, goals are redefined, if necessary, based on changing contexts.

There are three stages to an action evaluation. Stage one involves articulating success, developing common definitions within the project team, and developing subsequent action plans. Stage two sees the implementation of those action plans and the adjustment and monitoring of definitions and actions. By reflecting on the project experience to date, the team re-fines goals and develops strategies for overcoming obstacles. Stage three

An evaluation approach provides the framework, philosophy, or style of an evaluation.

Action evaluation is an iterative goal-setting process facilitated by an evaluator throughout the life of a project.

involves asking questions about the intervention and whether it has met its goals based on the criteria for success developed by the team. This discussion also explores why the project did or did not reach its goals and how things could have been done differently.

Supporters of action evaluation maintain that projects using this approach become more successful by reaching consensus about what they seek to accomplish, why, and then how. Action evaluation is useful for long-term projects operating in highly dynamic conflict contexts that need to be nimble if they are going to effect change. It is also a useful process for those who wish to instill “reflective practice” into their work. In terms of results measurement, this approach is best used in conjunction with other evaluation approaches that focus more on the “classic” evaluation component of data collection and judgment.

Empowerment Evaluation

The empowerment evaluation approach is designed to help people help themselves and to improve their programs using a form of self-evaluation and reflection. It uses both qualitative and quantitative methods and can be applied to individuals and organizations (although it is best used on projects). In this approach, the intervention participants and staff jointly examine issues of concern, while an external evaluator performs the role of a coach or extra facilitator depending on the needs and capabilities of the participants.

The empowerment evaluation approach is designed to help people help themselves and to improve their programs using a form of self-evaluation and reflection.

The approach recognizes that contexts, populations, goals, and knowledge shift and so too must the evaluation in order for it to remain current to the context and project so that it continues to gather relevant and useful information. As such it seeks to become an ingrained process within the intervention rather than a step within the project cycle.

There are five main facets to an evaluation that follows this approach: training, facilitation, advocacy, illumination, and liberation. The facets illustrate the experience of many who use this approach. In some cases, these facets may be treated like steps, though this is not their intended use. The training facet involves the staff and participants of an intervention being trained in how to conduct a self-evaluation. The training seeks to demystify evaluation and to help organizations internalize the principles of evaluation. The second facet, facilitation, sees the evaluator take the role of a coach whose task is to provide useful information and keep the evaluation on track. The evaluator also may play a significant part in the evaluation design process to ensure it meets the needs of the organization.

Advocacy, the third facet, sees evaluators acting as advocates for the intervention. This can include writing opinion articles for the newspaper, speaking to elected officials, or disseminating results to decision

makers. Of course, this assumes that the evaluation showed the program was of merit. The fourth facet, illumination, refers to the process of identifying new opportunities or connections for the project. It is often the combination of the previous four facets that allow for the fifth facet, liberation, to occur. In this facet, individuals involved are freed from pre-existing roles or ideas and new conceptualizations are developed for the project.

The benefits of this evaluation approach come to the organization that wants to incorporate monitoring and evaluation into its programming, but whose staff view evaluation as a foreign concept and question - if not reject - its usefulness. The costs come in terms of additional staff time, the ongoing presence of the external evaluator, and, potentially, in credibility – though that depends on the intended audience. If a more traditional donor is the primary audience, and the purpose is to prove progress toward results, this may not be the best model. In this case, if capacity building is still highly important, increasing the involvement of the external coach may be helpful. Empowerment evaluation advocates would maintain that this approach produces highly credible results because bias is minimized due to the broad range of stakeholders involved who then serve as a check on individual members' biases or agendas.

To capitalize on this approach fully, integrate it during the design stage so that the approach instructs the monitoring data collection, baseline as well as the evaluation. Using this approach does not exclude more traditional evaluation methods and, in fact, can be combined with them since the empowerment evaluation activities can provide a rich source of data.

Goal-Free Evaluation

Goal-free evaluation focuses on the actual results of a program rather than verifying achievement of the intended results. The goal-free approach evaluates a program with no knowledge or exposure to the predetermined goals and objectives of the program. The evaluator concentrates on what has actually happened as a result of the program rather than on the specific results intended by the program team. The results identified are then compared to the needs of the affected population to determine if the program was effective.

To undertake this approach, the evaluator has minimal contact with the program staff and intentionally avoids becoming familiar with the goals and objectives of the program. The data collection effort is open-ended and is intended to seek out all effects, positive and negative. Questions such as, "Have any changes in your community occurred over the past six months and why?" are used instead of, "After participating in the training, what changes occurred as a result in your community?"

Goal-free evaluation focuses on the actual results of a program rather than verifying achievement of the intended results.

Self-evaluation is an internally led review process that uses the same skills, standards, and techniques as all evaluations.

It is maintained that goal-free evaluation minimizes bias in the evaluation process because it is not based on the project logic of the program team. It is argued that, if the project logic is flawed, significant and valuable changes that have resulted from a project may be missed in a goal-based evaluation since it is an approach that seeks to determine if the pre-set goals have been accomplished. Moreover, informing the evaluator of the goals can consciously or unconsciously limit their perceptions consequently missing important information.

Goal-free evaluation is generally more costly than an approach using a goal-based approach because the evaluation team must inquire into a far broader set of issues with a wider range of stakeholders. On the other hand, the goal-free approach offers a true accounting of what difference a project has made because it is not dependent on the design staff being entirely accurate in their project logic. This approach is particularly effective in situations where the project logic has not been articulated or even considered as well as in innovative pilot projects where the project logic is based upon an as-yet-untested hypothesis not yet tested.

Self-Evaluation

Self-evaluation is an internally led review process that uses the same skills, standards, and techniques as all evaluations. The evaluator and those evaluated are identical. This approach has three core objectives: becoming aware of the complete picture, learning from experience, and adapting.

Self-evaluation can be done for individuals, teams, institutions, and projects. It is best for a group to initiate it, although it should be led by an individual. Qualitative and quantitative methods can be used with this approach. In addition to the methods listed in the Methods chapter on page 207, self-examination and reflection are also key sources of data.

The various self-evaluation approaches each involve different steps. Generally speaking, however, one starts with identifying the person who will be responsible for coordinating the process. Since all participants in a self-evaluation should have equal roles, care should be taken to avoid mirroring the organization's hierarchy in selecting a coordinator. In other words, it is not necessary for the self-evaluation leader to be the team leader for a team evaluation or for the organization president to spearhead an institutional evaluation. Rules of engagement or ground rules should be established for the participants in the evaluation, such as voluntary participation, freedom of speech, respect for others, agreement on the results, and dissemination of the results.

The next step is for those involved to develop a common understanding of the current situation. The key to this is to look back at the project's history. Be sure to allow all perceptions and viewpoints to be articulated so as to understand the full picture of the present situation. Analysis of the state of that situation comes next (i.e., where we are versus where we would like to be). As with all evaluations, this type of analysis requires the development of standards against which the work can be assessed. Setting norms or standards is essential to creating a truly effective process yet this is a step that is often missed in self-evaluations.

Self-evaluation differs from everyday or spontaneous reflections on the quality of the project's work and the difference that work makes because the latter often lack the critical distance necessary to gain a deeper and more accurate reflection of the mechanics of the situation. If done properly, self-evaluation provides a platform or conceptual basis that enables those involved to achieve the necessary distance.

At this point, discussions should be held on the basis of the information gathered and the standards set to seek agreement about the state of the current situation. This agreement will be the foundation for the project's new orientation that will result from the self-evaluation. Once agreement is reached, the next stage is to turn talk into action by incorporating what was learned into the new programming.

Those who use self-evaluation reap the benefits of an internally led assessment process in which staff integrate the value of questioning, setting goals, and assessing progress into daily practice. On the other hand, self-evaluation can cause internal conflict that, if not handled well, may continue to disrupt the working environment. It also requires substantial staff time to implement the process, and the process may lead to the collective bias of those involved, which can influence the conclusions. For teams and projects, self-evaluation is often a useful complement to monitoring systems in which a self-evaluation exercise is one element.

Theory-Based Evaluation

Theory-based evaluation focuses on why and how changes occur in programs. This approach focuses on the "black box" of programs, which is the space between the actual input and the expected results of the project. That space is called the black box because, in the effort to address highly complex and urgent social issues, program designers often gloss over how and why their intervention will address the issue at hand. This approach seeks to identify, articulate, explain, and then test the transformation process between input and results.

This transformation process is commonly referred to as a theory or theory of change, which can be broadly defined as a set of beliefs about how change happens. See page 14 in the chapter on Understanding Change

Theory-based evaluation focuses on why and how changes occur in programs.

for more information about theories of change for peacebuilding. The theory-based evaluation approach is based on the assumption that social change programs select and structure their interventions on the basis of some articulable rationale. Advocates for this approach maintain that, by purposefully identifying the assumptions on which programs are based, the design, implementation, and subsequent utilization of the evaluation will be better.

The theory underlying the program is located at the centre of the evaluation design. The first step in the evaluation, then, is to make the theory explicit, which may involve a process in and of itself if the theory has not previously been articulated or if there are conflicting ideas as to what underpins or creates change as a result of the program. All possible theories must be brought to light. The project team then selects which of the theories will be tested in the evaluation. Once the theory is articulated or selected, the next steps in the evaluation are built around it.

The theory-based approach can provide a project team or organization with an important facilitated process to unearth the “why” behind the projects that they implement. Understanding the rationale or theory upon which projects are based is essential for advancing our thinking as a field. Additionally, this approach allows us to test if the intervention was ineffective because of poor implementation or because the theory was flawed. For organizations lacking explicit theories of change, a theory-based approach will require some up-front work with the full team. Consequently, both the evaluation team and the project staff may need to allocate more time to the evaluation.

“Utilization-focused evaluation begins with the premise that evaluations should be judged by their utility and actual use...”

Utilization-Focused Evaluation

According to Michael Patton, the author of this approach, “utilization-focused evaluation begins with the premise that evaluations should be judged by their utility and actual use; therefore, evaluators should facilitate the evaluation process and design any evaluation with careful consideration of how everything that is done, from beginning to end, will affect use.” Patton’s approach offers a process by which the evaluator and the intended users can collaboratively reach agreement on how the evaluation will meet the needs of the project team. In this manner, the evaluator works with the primary intended user to facilitate the users’ decisionmaking on use, method, model, and theory.

The evaluator begins the process by spending significant time with the project team to establish a common definition of evaluation. She/he ensures that the team has a clear understanding of what is involved in an evaluation process. While developing this common understanding, the team is encouraged, through a variety of exercises and games, to express any fears or concerns about the evaluation. The team is also

invited to consider the incentives and barriers to engaging honestly in the evaluative process. As these issues are raised, the evaluator discusses them openly so that they can be handled constructively and jointly.

Utilization-focused evaluation is based on the premise that an evaluation will not be utilized if the user does not have ownership of the evaluation from the start. Ownership can be fostered through an increased understanding of the benefits of evaluation. It can also be engendered through confidence that the process will support and help rather than challenge and criticize. The first stage of the utilization-focused approach is intended to build ownership.

The evaluator then works with the project team to identify exactly who will be the primary user of the evaluation. Again, this is a facilitated process in which the evaluator works collaboratively with the project to jointly reach an answer. Once this has been identified, the process moves toward identifying the actual intended use by generating questions. There are five criteria used to develop utilization-focused evaluation questions. First, data can be collected that directly answers the question. Second, there is more than one possible answer to the question. Third, primary intended users care about the answer and want information to help answer it. Fourth, primary users want the answer to the question for themselves, rather than just giving it to a second party such as a donor or the press. Fifth, the users can describe how the answer may change actions in the future.

There are no predetermined methods for this approach and it blends very well with other approaches. It is a good choice when there is a high degree of resistance to evaluation among the project team. The utilization-focused approach is also useful in situations in which there appears to be a large number of evaluation objectives that the team is unable to limit or decrease. While valuable for both formative and summative evaluations, this approach requires that time be allocated for the preparation process described. This is the approach most heavily drawn upon in this manual.

Evaluation Approaches: Pro-Con Framework

| Approach | Pros | Cons |
|----------------------|--|---|
| Action Evaluation | <ul style="list-style-type: none"> Facilitates project adaptation to changing environments Ensures goal agreement within team | <ul style="list-style-type: none"> Primary focus is on the design element of the process and less on the gathering of evidence to prove results |
| Empowerment | <ul style="list-style-type: none"> Builds staff M&E capacity Combines internal and external expertise and perspective | <ul style="list-style-type: none"> May be deemed less credible due to internal contribution by staff Not all evaluators will be comfortable playing an advocacy role |
| Goal-Free | <ul style="list-style-type: none"> Captures unintended negative and positive effects Limits bias of project team and evaluator | <ul style="list-style-type: none"> Requires more time and funding than other approaches Results may not be sufficiently concrete to act upon |
| Self-Evaluation | <ul style="list-style-type: none"> If resources are allocated to skills development, this approach builds internal M&E capacity Owned and implemented by staff Should reflect the real needs and questions of the team | <ul style="list-style-type: none"> May be deemed less credible by external audiences Has potential for bias No outside perspective to challenge assumptions Susceptible to internal political pressure |
| Theory-Based | <ul style="list-style-type: none"> Articulates assumptions that underpin the work Uncovers differing views on theories of change or the “why” Allows the flaws in theory to be distinguished from poor implementation | <ul style="list-style-type: none"> May have a heavy up-front time commitment if the theory of change has not been articulated Innovative programs may not have the theory accurate yet. The focus on flawed theory may overlook changes that the program is effecting |
| Utilization- Focused | <ul style="list-style-type: none"> Increased likelihood that evaluation results will effect change in the project or organization Decreases emotional barriers to the idea of results and measurements that will last beyond this evaluation | <ul style="list-style-type: none"> Requires more time at beginning of process May backfire if the evaluator does not have the necessary facilitation skills |

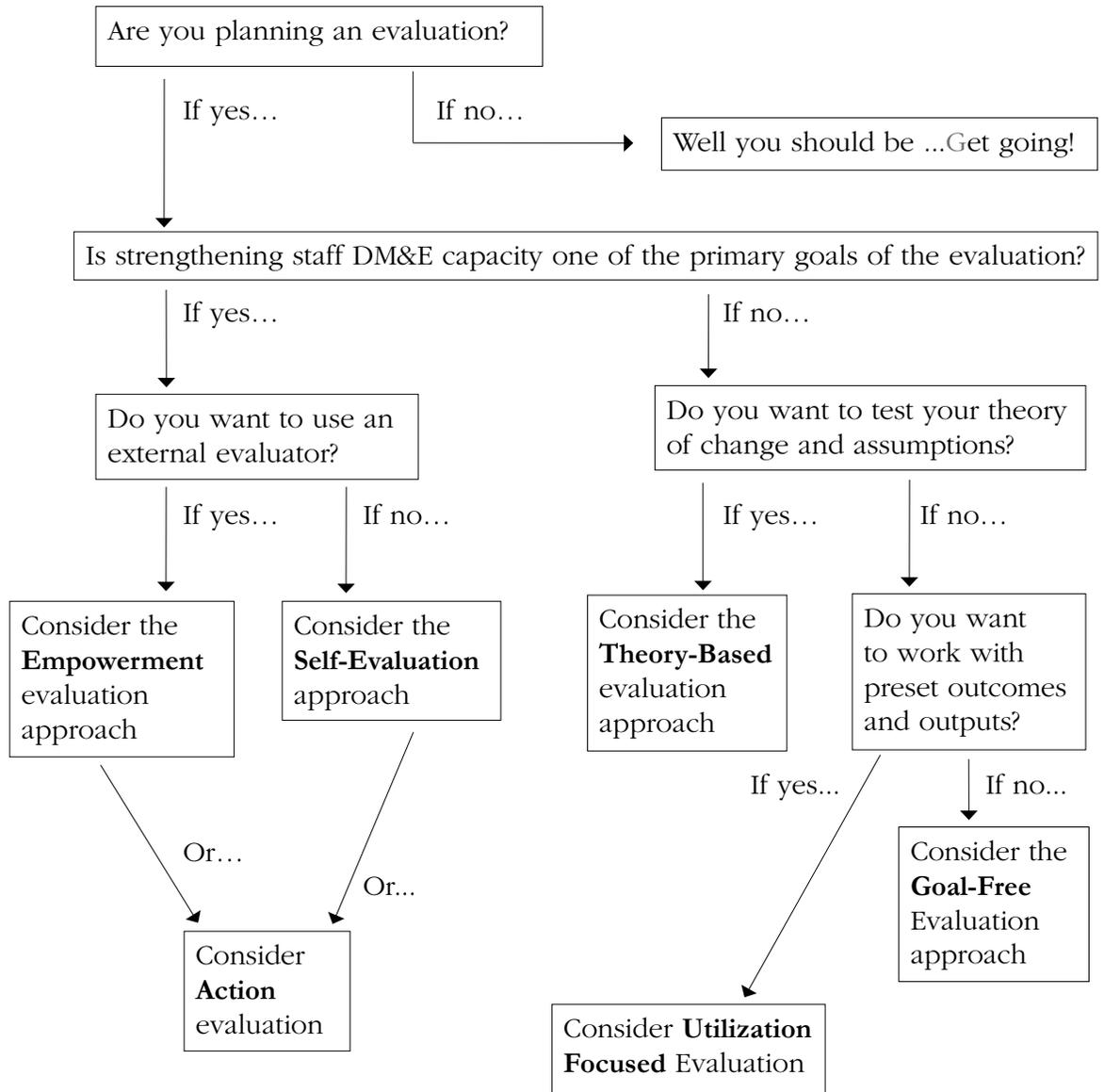
What should I consider when selecting the approach?

When considering approaches, reflect on the evaluation objectives since they should be the point of first departure for this decision. For instance, if implementation process appraisal is the evaluation objective and the team wants recommendations for improvement, the utilization-focused approach best meets those specific needs. On the other hand, if the same team feels that it has not explicitly examined its underlying assumptions behind the intervention, a theory-based approach is a natural choice.

Sometimes there will be natural and direct fits between the evaluation objective, the scope, and the approach. When that does not occur, however, feel free to mix and match among the approaches to pick the components that best meet your needs. For instance, one could blend the evaluator coach found in empowerment evaluation with the theory-based articulation and testing of the theory of change. If one uses a blend of approaches, it is advisable to discuss this with the evaluation team before making a final decision. This allows the evaluator to advise the team on the feasibility and implications for cost and time of different combinations. Also, keep in mind that any increase in the number of components blended implies the need for a broader skill set from the evaluators, which may make it more challenging to recruit qualified individuals.

To help with decisionmaking, a decision flowchart is located on page 123. For the sake of simplicity, the flowchart only asks core questions, although this is not meant to diminish the importance of other variables.

Evaluation Approaches Decision Flowchart



Example

IDP-Host community conflict

Part 5: We know that the project team is planning an evaluation; therefore, we can proceed along the “yes” arrow. Strengthening DM&E capacity is not a primary focus of this evaluation process, so we can follow the “no” arrow. Although the team is interested in the theory and underlying assumptions of the project, the team members do not think they will make changes at this stage of the project. Thus, they do not want to test theories and assumptions. They do wish, however, to establish whether or not their objectives are being met as originally drafted. Consequently, the team will select a utilization-focused approach to this evaluation.

VI. DECISION: EVALUATION SCOPE

What is the scope of the evaluation?

The notion of scope encompasses two inter-related aspects: geographic coverage and the degree to which conclusions will be generalized. Each of these should be considered for every evaluation objective. Sometimes the geographic coverage of the evaluation is the same one as the intervention itself. Consider a project that seeks to change the relationship from a negatively charged one to a cooperative one between the Morley tribe of First Nations people and the neighboring rural communities in a municipality in Alberta, Canada. The geographic coverage of the evaluation could be the entire municipality, which would be the same coverage as the project.

If there was a specific area or location within the municipality that seemed to be either progressing exceptionally well or regressing, it could be singled out as the sole focus of the evaluation. Here, the evaluation geographic scope would be less than the project geographic scope. Conversely, that scope could be far greater if the team wanted to understand whether or not there were any transfer or ripple effects beyond its immediate working location.

The second aspect of evaluation scope is whether or not the conclusions need to be generalized to include the entire population, be it a tribe, village, community or country. To generalize in this way requires that there be “enough” data sources to draw a conclusion that could represent that entire population. Called statistically significant conclusions or generalizations, what constitutes “enough” data sources is determined by the size of the population. This type of information can be extremely valuable, but acquiring it has significant cost and time implications.

Take, for instance, an intervention in the form of a television edu-drama aimed at young men (ages 15-25) and broadcast across Palestine that challenges Palestinian attitudes on obstacles to the Palestinian-Israeli peace process. The project team wants information from the evaluation that is representative of all young men in Palestine. This will require evaluators to gather “enough” data to draw generalized conclusions that can speak for the target population. If the evaluators do not gather “enough” data, they can only draw conclusions for the people who participated in the evaluation.

In this case, the geographic scope and the project scope are the same. Generalized conclusions are not tied to national geographic coverage. One can request generalized conclusions for any group as long as the group is well defined.

Where conclusions cannot be generalized because of the small number of people involved in the evaluation, it is important to indicate that the conclusions cannot be applied to the wider group.

On the other hand, if the project team felt that it would be more useful to learn about the views of those who live in the most violent area or on the border, the geographic scope of the evaluation would be smaller than the scope of the intervention. The evaluators would only go to those areas in the country that were violent or near the border. Within this more limited scope, the project team may still want conclusions that apply to the entire population of those defined areas (e.g., a border region). Alternatively, the ability to generalize may not be relevant to the evaluation. If, for example, there is an evaluation objective of strategic alignment, it may not be necessary to have generalized conclusions to inform the project team about that objective.

It is often not possible, due to time, cost or security limitations, to gather enough data to generalize conclusions for large populations, such as the citizenship of an entire country. Where conclusions cannot be generalized because of the small number of people involved in the evaluation, it is important to indicate that the conclusions cannot be applied to the wider group. As the need to generalize expands to larger groupings, so too does the cost and time required. In addition, the evaluators will need to have the expertise suitable to ensure that the methods used are credible. It will also require a careful assessment of security measures to ensure that the evaluation team can access a broad enough range of locations to gather the necessary data.

Example

IDP-Host Community Conflict

Part 6: The NGO has decided that the geographic scope for all objectives of the evaluation is the locations of the communities that have IDPs. At this stage, the project team’s number one priority is the immediate target group. The type of information the team seeks is, therefore, best given from this group. In terms of the process implementation appraisal objective, the project team does not feel it is necessary to have generalized conclusions. For the outcome identification objective, however, the team does want to have conclusions that are representative of the entire target population. This is important for the project team because its members want to be certain that the changes are occurring throughout their target group. For this project, the target population is the IDP and host community leaders for each worksite.

VII. DECISION: EVALUATOR QUALIFICATIONS

Who should conduct the evaluation?

At this stage, it is necessary to have a broad understanding of the type of person needed to conduct the evaluation. Of course, this does not mean either writing down the required qualifications in detail or actually selecting the person. This decision about the type of evaluator needed informs the budget structure and provides an indicator of the difficulty there will be in recruiting a qualified individual. The latter point is important if a highly specialized person is required, because recruitment will need to start earlier than normal.

- Will the evaluators be internal to the organization or external?
- What type of experience is required?
- How many evaluators are needed?
- Will the evaluators be local hires or recruited internationally?
- Will they need translation services?

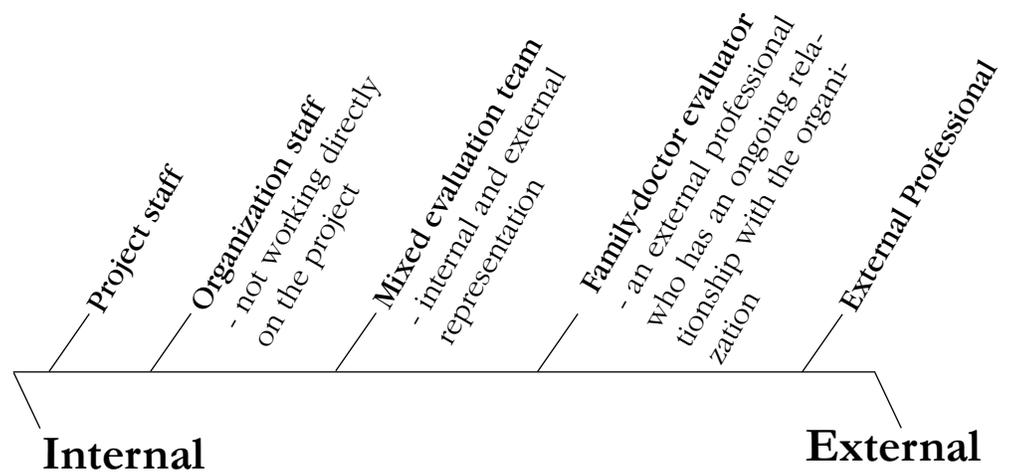
There are many more decisions to be made regarding specific qualifications but those can be left to page 140 of the chapter 9, “What do the Terms of Reference contain?”

Will the evaluators be internal to the organization or external?

The first step is to determine whether the evaluators will be internal to the organization (staff) or external (professional). The internal-external categorization is dependent on the relationship of the evaluator with the organization; it does not refer to where the evaluation originated. The internal-external notion is actually a continuum,²⁰ with project staff members doing their own evaluation (i.e., self-evaluation) on one end and an external professional who has never worked with the organization on the other end.

The internal-external categorization is dependent on the relationship of the evaluator with the organization; it does not refer to where the evaluation originated.

²⁰ Modified from Church and Shouldice, *Part II*.



The approach and scope desired for the evaluation may set capacity requirements that exclude many internal staff.

In deciding which of these relationships is best for the evaluation, there are several considerations to keep in mind, not the least of which are practical considerations in terms of costs, capacity, and time. The internal end of the spectrum is often less expensive for the organization, but it does require a level of skill that is not commonly available internally. In terms of time, staff members would need to be able to leave their regular duties for a period of time to undertake some parts of the evaluation. In addition, for evaluations in identity-based conflict areas, internal staff may not be able to enter certain communities or obtain factual answers from members of the “other” community or identity group.

One of the benefits of using internal staff members is that the learning (such as evaluation skills and experience) and the programmatic conclusions that may be applied to their own projects stay within the organization. Furthermore, staff members are familiar with the organizational history, approach, and assumptions, which allows for a shorter learning curve in this area.

The external end of the spectrum, on the other hand, has higher associated costs and requires more time to recruit and orient the team. That said, external evaluators can be highly effective in challenging implicit assumptions and organizational norms that have become standard practice, and they offer a broader range of expertise, which can be of great benefit. In addition, externals are generally deemed to be more objective and removed than internal evaluators, which can give the final product greater credibility.

Take for example, the evaluation objective strategic alignment. Someone internal to the organization but not on the project staff may be best placed for this assignment because she/he has a deep understanding of the organization’s mission and principles. Conversely, if the evaluation objective is to conduct outcome identification, an external evaluator may be better suited for such an assignment because she/he will not be bound to the assumptions behind the desired changes.

The approach and scope desired for the evaluation may set capacity requirements that exclude many internal staff. This is an important point

since evaluation conclusions are only as good as the implementation of the data collection techniques. If internal staff do not have the necessary skills, and it is deemed valuable to build in-house capacity in this area, then creating a mixed team might be useful. If this approach is adopted, extra time should be factored into the evaluation for the external professional to coach the staff person in her/his role.

Example

IDP-Host community conflict

Part 7: A mixed team (staff and external professionals) would be a good choice for this evaluation. The implementation process appraisal aspect would benefit from a staff person from outside the project team who is knowledgeable about the organization’s model and techniques for this type of work. On the other hand, an external evaluation professional would offer new insights and challenge assumptions on this aspect of the evaluation. The output and outcome identification objectives would capitalize on the external evaluator’s experience in data collection and would benefit from the credibility lent by that external professional. Since the outcome identification objective also requires generalized conclusions, the external evaluator will need to have expertise in this subject area. The staff team member would benefit from the experience as they would build her/his skills in these areas.

What type of experience is required?

At this point in the evaluation preparation, there are three factors that assist in determining the type of experience required:

- Evaluation Objective
- Approach and Scope
- Evaluator’s Role

If, for instance, the organization has identified appropriateness consideration as an evaluation objective, the evaluator should have knowledge of peacebuilding theories of change and an in-depth understanding of the conflict context. If the second evaluation objective was outcome identification and the organization required generalizable conclusions (scope), the evaluator hired would need to have statistical expertise and quantitative research experience. Finally, if the evaluation required a learning facilitator, the evaluator would need facilitation skills and, ideally, conflict transformation program design experience.

How many evaluators are needed?

The number of evaluators needed is directly related to the amount of work involved, how much time is available, and the variety of skills needed to accomplish the evaluation. One or two people could do an evaluation that was national in scope in Burundi if it were limited to one evaluation objective and nine months in which to conduct it. The same evaluation, if needed in a month, might require a team of five. In a country the size of Indonesia, the same evaluation might need double the staff of the previous example. If the evaluation is complex, a broader set of skills are often required, which may demand a team approach in order to have all skills represented.

Will the evaluators be local hires or recruited internationally?

This question depends on the location of the program and the availability of local professionals. Local professionals understand the culture and history of the conflict, yet they also may identify too closely with it. Outside North America and Europe, local professionals are often a more cost-effective option for an NGO. Evaluation as a profession is still relatively new, however, which means the pool of qualified candidates is more limited, though this is changing rapidly.

Where possible, utilizing a mix of local and international evaluators can offer an effective blend of cultural sensitivity and external “neutrality” while also maximizing the skill set on the team.

Example

IDP-Host Community Conflict

Part 8: This evaluation team will consist of three individuals. One of them will come from the NGO staff and will be familiar with the model and techniques used elsewhere in the organization. This person will focus on implementation process appraisal. Ideally, the person would be located in the Sudan office, but would not work on this project directly. Another will be an external evaluator, recruited internationally but with experience in the region, who will be the team leader and who will provide direction and input to both evaluation objectives. The third individual will also be an external evaluator, recruited from Sudan, who will focus on outcome identification. This person’s role will be essential in ensuring the cultural sensitivity and accuracy of this aspect of the evaluation. Three evaluators are required for this project due to the added workload that comes with providing generalized conclusions (because far more people will need to be included in the data collection). If that had not been the case, two evaluators would suffice.

Will translation be needed?

The need for the evaluator to speak the local language depends upon the people from whom the evaluator will be predominately accessing information and the data collection tools (e.g., surveys and interviews) that will be utilized.

If the evaluation approach includes extensive interaction with local communities, an ideal scenario would be for the evaluator to speak the local language(s). This facilitates greater access to the community, better mobility, and a more nuanced understanding of expressions and terms. If a qualified person cannot be found, the organization may wish to consider working via a translator(s).

The intricacies of successful translation are often not fully considered, which can negatively affect the quality of the evaluation conclusions. One should consider the individual's experience as a translator as well as her/his background as it relates to the conflict. Nuance and choice of language are critical to most qualitative evaluations and, as such, can be subconsciously altered by an amateur translator who has strong opinions or biases regarding the subject matter. Furthermore, if the work is in an identity-based conflict, the translator's identity must also be considered. In Kosovo, for instance, one would find it difficult to obtain honest responses from the Kosovar Serbs if working with a Kosovar Albanian translator, despite the quality of the translation.

A translator can also act as a gatekeeper for a community. This can be positive in that it can facilitate unprecedented access for the evaluator in some cases. On the other hand, the translator may feel obligated to represent the community in the best light possible and, as such, direct the evaluation team to individuals who paint a particular portrait of the situation.

One common approach, when there are financial or capacity restrictions, is to assign a project staff person to be the translator. This has the benefits of being both cost effective and providing the evaluator with a ready source of project and community information. However, it can adversely affect the interviews if project participants do not wish to say something negative in front of the project staff. The politics of evaluation can become very clear in this case, if, for instance, the project participants feel that ongoing participation could be contingent on their answers.

Using an internal translator is not a recommended strategy; however, if financial or capacity restrictions require this approach, the evaluator should try to interview some people without using the translator by using a common language such as French or English to see if different answers are given without the translator's presence. Where possible, have

The intricacies of successful translation are often not fully considered, which can negatively affect the quality of the evaluation conclusions

A translator can also act as a gatekeeper for a community.

staff members involved with participants that they do not know or work directly with so that the personal relationships do not restrict what the individuals say.

Undoubtedly there are conflict areas where access to individuals or key groups cannot be obtained without the presence of project staff who have built trust and relationships. In these cases, the ability to access key people may outweigh the potential changes in answers of having project staff present. That said, the evaluator should either identify a means to control for this or, at a minimum, acknowledge it in the report.

VIII. DECISION: TIMING

When will the evaluation take place?

With the decision regarding formative, summative or impact evaluation type already made, the question of the evaluation date becomes quite straightforward in many cases. That said, evaluation planners should review the proposed timing with a conflict lens to ensure that the evaluation will not take place during a time of predictable unrest or jubilation since this could affect the evaluation conclusions. Memorial days, election, and anniversaries of peace accords are examples of predictable events that could either spark tension or create unsustainable optimism.

In Northern Ireland, for example, it would be unwise to plan an evaluation during the summer months due to the tension caused by the annual marching season. These events may not only change the attitudes and behaviors at that time, they can affect the ability to gain access to places due to violence in extreme cases or people's willingness to speak with outsiders.

There are also some practical concerns to take into account:

- Will key staff be in the country and available to the evaluators during the proposed evaluation period?
- Is there a strategic planning date that the evaluation findings should inform?
- Are there any weather conditions, such as a rainy season or extreme snow, that might limit the evaluator's ability to travel to different parts of the country?

- Will the program participants be accessible? For instance, it would be difficult to conduct a peace education program in schools during the summer months when children are out of school.
- Does this conflict with any major events or deadlines within the project being evaluated?

Example

IDP-Host Community Conflict

Part 9: The evaluation is formative so it needs to happen broadly in the middle of the project, which offers a window of September-February. A careful look at the calendar suggested November as the best month. That date avoids the local elections scheduled for mid-September and the rainy season which comes in late spring. In addition, all staff will be in the office in November and the next donor report is due February 1st.

**IX. DECISION:
BUDGET**

What will the evaluation cost?

Budgeting is the last decision at this stage. All the decisions preceding this one affect the budget. Is the evaluation national in scope or limited to one or two communities? To fulfill the evaluation objectives, does it require a multi-faceted evaluation team or one individual? Is the evaluator's role to engage with the project team and build capacity along the way or to implement the evaluation? Once clarity has been reached on these decisions, it is time to create the budget estimate so it can be included in the project proposal.

There is an argument that one should start with the maximum budget predetermined so that the preparation is realistic from the beginning. Though seemingly practical, this approach limits creativity and often a true unearthing of the needs and learning desired from the experience.

The Evaluation Budget Worksheet on page 133 details those line items commonly found in an evaluation budget. Not all of these line items will apply every time, particularly because different data collection methods have different costs associated with each. One can also use the rule of thumb that the baseline, monitoring, and evaluation costs will constitute 5-10% of the project budget.

One can also use the rule of thumb that the baseline, monitoring, and evaluation costs will constitute 5-10% of the project budget.

Although it is not necessary to select the methods in the evaluation preparation stage, this worksheet includes an illustration of those costs. It includes expenses typical of a hard-copy questionnaire disseminated by post and focus groups. An evaluation that uses surveyors, for example, would need to include the fee to hire the surveyors plus the costs associated with training them such as rental of the training venue, meals, and equipment (clipboards, ID cards, flashlights, etc.). To do the budget estimate, one would insert a broad estimate in that portion of the worksheet based on what is already known about the approach and scope.

Budget Worksheet

| Common Budget Line | Description/Comment |
|--|--|
| Evaluator Day Rate (fee) | This is the fee paid to the evaluator. It is most commonly determined on a daily rate, though it could be computed as a lump sum amount. In 2005, daily rates of Western European/North American NGO evaluators ranged from US\$250-700. |
| Return Economy Airfare to Home Country | Pre-booking and staying over Saturday nights are effective means to keep this cost low. |
| Visa Cost | Often forgotten, this can be as high as US\$300, so it is worthwhile to include this cost. |
| Immunization Cost | Not all countries require shots; however, most African countries do and the cost can add up. |
| Evaluator In-Country Travel | This line item can include renting a vehicle, hiring a driver, domestic flights or even boats. |
| Evaluator Per Diem | The per diem is the amount given to cover the evaluator's daily food and incidental expenses and, in some cases, hotel as well. Since different countries and organizations have different norms for what is included in the per diem, detailing in the contract what your per diem includes will reduce the chance of a misunderstanding.. The U.S. government posts its official per diem rates online at: http://www.state.gov/m/a/als/prdm/2004/29997.htm |
| Evaluator Hotel | Number of nights in a hotel multiplied by the price per night, if not included in the per diem. |
| Translator Day Rate | This is the fee for each day. The average price range varies from country to country. Note that if language differs between parties to the conflict, consider whether two translators will be required. |

Budget Worksheet

| Common Budget Line | Description/Comment |
|------------------------------------|---|
| Translator In-Country Travel | Be sure not to double count. If the cost of a rented vehicle has been included already there is no additional cost for the translator. A domestic airplane seat, however, would need to be included. |
| Translator Per Diem | Same concept as the evaluator per diem. If the translator lives nearby, she/he does not need the full per diem since she/he will be returning home at the end of each day. However, the translator is often expected to eat lunch and even dinner with the team because those are working meals. In such cases, it is equitable to provide the translator with a portion of a per diem. |
| Translator Hotel | Number of nights in a hotel multiplied by the price per night, if not included in the per diem. |
| Data Collection Tool Costs: | Questionnaire & Focus Groups |
| Questionnaire Translation | The cost of translating the questionnaire into the local language. |
| Photocopying | Cost of copying the questionnaire. |
| Postal Charges | Costs include envelopes, postal charges, and providing stamped return envelopes. |
| Venue Rental | Focus groups often need a venue that is beyond the capacity of the local NGO to provide. Note that in conflict settings one community may not be able to travel to an NGO office that is located in the “other” community. In other contexts, focus groups may be hosted outdoors at no cost. |
| Food or Beverages | Focus groups are often supplied with a beverage, at a minimum. |
| Data Entry | For large questionnaires, sometimes data entry companies or individuals are hired to enter the data for statistical analysis. |
| Participants’ Time Compensation | In some places, the time used to attend a focus group is time taken away from earning the next meal. In these cases, providing some form of compensation in the way of a meal or transportation costs should be considered. |
| Administrative Expenses | This includes all basic administrative costs such as conference calls with the project team and evaluation team or shipping final bound versions of the evaluation to offices. |
| Total | |

How long does this process take?

The length of the evaluation preparation process depends on the complexity of the project, the number of decision-makers on the project team, the familiarity of the project team with evaluation concepts, and the understanding they have of the relationship between concepts as depicted in the Evaluation Preparation Decision Flowchart. As the complexity of project and team increases, so too does the time needed to work through this flowchart.

Assuming that there is a reasonably detailed outline of what the project will entail, the evaluation preparation process for those teams on which all actors are familiar with the concepts - and the interconnectedness of those concepts - can be handled in one session. The length of the session will depend on the issues referenced above.

If the actors are not familiar with the concepts, they should, ideally, work with an evaluation facilitator to remedy their lack of knowledge, which will decrease potential frustration and time needed for explanation in the future. In this case, it normally takes two separate sessions – although sometimes three – to first explain the concepts and then the pros and cons of the different options. In addition, newcomers to these ideas should be offered the chance to reflect on their decisions between sessions to ensure that the evaluation truly meets their needs.

Keep in mind, however, that the time spent in the project design stage is time saved later when the proposal writer doesn't have to struggle to fill in the evaluation section or during the project implementation, as the team starts to consider the evaluation.

Further Reading

Johanna Birckmayer and Carol Weiss, *Theory-Based Evaluation in Practice: What Do We Learn?*, Evaluation Review, SAGE Publications, Vol. 24, No 4, August 2000.

D. Fetterman, S. Kaftarian and A. Wandersman (eds.), *Empowerment Evaluation, Knowledge and Tools for Self-Assessment and Accountability*, SAGE Publications, 1996.

Michael Quinn Patton, *Utilization-Focused Evaluation: The New Century Text, 3rd Edition*, SAGE Publications, 1996.

Jay Rothman, “Action Evaluation” in *Beyond Intractability*, Guy Burgess and Heidi Burgess (eds.), October 2003. http://www.beyondintractability.org/m/action_evaluation.jsp.

Michael Scriven, *Pros and Cons about Goal-Free Evaluation* in Evaluation in Education, W.J. Popham (ed.), McCutchan Publishing, 1974, pp. 34-67.

Swiss Agency for Development and Cooperation, *Manual on Self-Evaluation*, February 1996. http://www.deza.admin.ch/ressources/deza_product_e_118.pdf.



**STAGE
TWO**

EVALUATION MANAGEMENT

This chapter contains:

1. Developing the Terms of Reference
2. The Evaluation Plan
3. Frequently Asked Questions about Working with External Evaluators
4. Strategies for Overcoming Common Evaluation Pitfalls

INTRODUCTION

Evaluation Management involves the implementation of decisions made in the preparation stage. The process begins with the creation of the terms of reference (TOR), which generally takes place four to six months prior to the evaluation date. Developed by the project team, the TOR is essentially a guide to the evaluation and, as such, needs to be well thought through.

After completing the TOR, the next step in the evaluation management stage is to develop the evaluation plan. This can be done by the project team or by the evaluators and always should have input from both parties. As the second step in the evaluation management process, the evaluation plan operationalizes the decisions made in the TOR. This requires additional decisions related to the means of verification, data sources and targets, location of data collection, conflict considerations, means of analysis, and timing.

Working with external evaluators also requires some thought to ensure a successful experience. This chapter walks through a number of the most frequently asked questions about external evaluators starting with the recruitment process through to what to do with the evaluator’s primary data after the evaluation is completed.

The Managing Evaluations chapter concludes by offering some strategies for overcoming common evaluation pitfalls.

Evaluation Management involves the implementation of decisions made in the preparation stage.

I. DEVELOPING THE TERMS OF REFERENCE

What are the terms of reference?

The first step in evaluation preparation is the development of the terms of reference (TOR). The TOR, also commonly called a scope of work, is a key part of the preparation stage. It is effectively a guide to the evaluation describing the objectives, deliverables, methods, activities and organization of the intended evaluation. The more preparation and thought that go into the terms of reference, the more likely the evaluation will be used to shape and inform decision making in the future.

The TOR is not only the “evaluation guide,” it is also commonly used as the basis for the contract with evaluators and as part of the recruiting materials for prospective consultants.

It is effectively a guide to the evaluation describing the objectives, deliverables, methods, activities and organization of the intended evaluation.

Further Reading

Johanna Birckmayer and Carol Weiss, *Theory-Based Evaluation in Practice: What Do We Learn?*, Evaluation Review, SAGE Publications, Vol. 24, No 4, August 2000.

D. Fetterman, S. Kaftarian and A. Wandersman (eds.), *Empowerment Evaluation, Knowledge and Tools for Self-Assessment and Accountability*, SAGE Publications, 1996.

Michael Quinn Patton, *Utilization-Focused Evaluation: The New Century Text, 3rd Edition*, SAGE Publications, 1996.

Jay Rothman, “Action Evaluation” in *Beyond Intractability*, Guy Burgess and Heidi Burgess (eds.), October 2003. http://www.beyondintractability.org/m/action_evaluation.jsp.

Michael Scriven, *Pros and Cons about Goal-Free Evaluation* in *Evaluation in Education*, W.J. Popham (ed.), McCutchan Publishing, 1974, pp. 34-67.

Swiss Agency for Development and Cooperation, *Manual on Self-Evaluation*, February 1996. http://www.deza.admin.ch/ressources/deza_product_e_118.pdf.



**STAGE
TWO**

EVALUATION MANAGEMENT

This chapter contains:

1. Developing the Terms of Reference
2. The Evaluation Plan
3. Frequently Asked Questions about Working with External Evaluators
4. Strategies for Overcoming Common Evaluation Pitfalls

INTRODUCTION

Evaluation Management involves the implementation of decisions made in the preparation stage. The process begins with the creation of the terms of reference (TOR), which generally takes place four to six months prior to the evaluation date. Developed by the project team, the TOR is essentially a guide to the evaluation and, as such, needs to be well thought through.

After completing the TOR, the next step in the evaluation management stage is to develop the evaluation plan. This can be done by the project team or by the evaluators and always should have input from both parties. As the second step in the evaluation management process, the evaluation plan operationalizes the decisions made in the TOR. This requires additional decisions related to the means of verification, data sources and targets, location of data collection, conflict considerations, means of analysis, and timing.

Working with external evaluators also requires some thought to ensure a successful experience. This chapter walks through a number of the most frequently asked questions about external evaluators starting with the recruitment process through to what to do with the evaluator's primary data after the evaluation is completed.

The Managing Evaluations chapter concludes by offering some strategies for overcoming common evaluation pitfalls.

I. DEVELOPING THE TERMS OF REFERENCE

What are the terms of reference?

The first step in evaluation preparation is the development of the terms of reference (TOR). The TOR, also commonly called a scope of work, is a key part of the preparation stage. It is effectively a guide to the evaluation describing the objectives, deliverables, methods, activities and organization of the intended evaluation. The more preparation and thought that go into the terms of reference, the more likely the evaluation will be used to shape and inform decision making in the future.

The TOR is not only the "evaluation guide," it is also commonly used as the basis for the contract with evaluators and as part of the recruiting materials for prospective consultants.

Evaluation Management involves the implementation of decisions made in the preparation stage.

It is effectively a guide to the evaluation describing the objectives, deliverables, methods, activities and organization of the intended evaluation.

When should the terms of reference be developed?

Evaluation management should begin four to six months before the evaluation is to begin. The longer the timeframe you have to develop the terms of reference, the less intense the work will be, which will make it easier to weave the tasks related to the evaluation into existing workloads. In addition, the more opportunity there is to reflect and make the most useful choices, the more time there will be to recruit the most qualified evaluation team.

Experienced evaluators are generally not available to work on an evaluation with less than two-three months notice. Therefore, the more time they have between when they receive notification about the evaluation and when the evaluation is expected to start, the greater the chance that the evaluator of choice will be available during the desired timeframe.

Who is involved in developing the terms of reference?

The key actors in the Evaluation Preparation decisions should meet to review their decisions to ensure that project needs are met. This generally includes the project team and the DM&E technical assistant. This is particularly helpful if new staff have joined the team after the design stage as they are able to add their input. This same group does not need to jointly manage the entire evaluation; however, they should be involved, at a minimum, in determining the Lines of Inquiry and in discussing the Evaluation Methodology. Their contribution can continue beyond this, although it may not be feasible in terms of balancing workloads.

At this point, an Evaluation Manager should be appointed to oversee and guide the evaluation from this time forward. Ideally, this person should not be part of the project to be evaluated. This is an important role that creates a champion for the process and distinguishes between those being evaluated and those who are responsible for the quality of the evaluation. The establishment of this position should in no way exclude the project team from the process; rather, provide a buffer between them and the evaluators.

Some of the decisions necessary to complete the TOR can often be better made in conjunction with input from the evaluation team. In this case, the organization might create a “draft” TOR for recruitment of evaluators who would then help complete the TOR once they have been hired. This is most commonly done for the evaluation methodology decisions. For instance, deciding how many sites of a multi-site project in Angola need to be visited may be made easier with the evaluators’ input.

Evaluation management should begin four to six months before the evaluation is to begin.

The key actors in the Evaluation Preparation decisions should meet to review their decisions to ensure that project needs are met.

It is important that the terms of reference be finalized and agreed upon by all parties before the evaluation team begins collecting data. Additional changes may be required as the process unfolds, which should be made in writing with the agreement of all parties involved.

What do the terms of reference contain?

The terms of reference can be organized in many different ways and the sections vary by agency. The most common sections are discussed below. What gets fed into each section of a TOR should be tailored to each new evaluation.

It is important that the terms of reference be finalized and agreed upon by all parties before the evaluation team begins collecting data.

Overview

This summarizes the key elements of the evaluation including:

- Evaluation type
- Dates
- Duration
- Intervention
- Country
- Request for applicants

Background

INTERVENTION SUMMARY: This is a maximum of two pages in length and can be supplemented by the completed management tool (logical framework or results framework). The organization can attach it as an appendix to this section or make it available upon request. This section covers key information on the project such as:

- Project goals
- Project objectives
- Current activities
- Location of activities
- Target audience
- Size or scale of project
- Duration of project

ORGANIZATION OVERVIEW: This offers key information about the organization including:

- Organizational mission
- Years in existence
- Size in staff, country offices or budget

| Terms of Reference Standard Sections |
|---|
| Overview |
| Background |
| Intervention Summary |
| Organization Overview |
| The Evaluation |
| Evaluation Goal |
| Evaluation Objectives |
| Lines of Inquiry |
| Audience |
| Evaluation Methods |
| Implementation Information |
| Evaluation Manager |
| Location |
| Deliverables |
| Duration and Working Days |
| Deadlines |
| Logistical Support |
| The Evaluation Team |
| Role of Evaluators |
| Evaluator's Responsibilities |
| Evaluator's Qualifications |
| Application Guidelines |
| Budget Guidelines |
| Contact Details |

- Length of time in country
- Types of programming offered

The Evaluation

EVALUATION GOAL: This indicates what is ultimately sought from the evaluation. This manual operates from the premise that the evaluation goal is “to improve peacebuilding programming practically and conceptually,” though other goals are possible.

EVALUATION OBJECTIVES: This lists the evaluation objectives. These represent what the project team has deemed they want to learn through the evaluation process.

LINES OF INQUIRY: The evaluation objectives set the theme(s) for the evaluation. To ensure that the exploration within that theme delivers useful information to the project team, lines of inquiry are added. Lines of inquiry are a series of questions developed to provide greater direction on what one wants to find out – facts, opinions, experiences, unintended effects, etc. They are often included in the terms of reference in combination with the evaluation objectives. The table on page 142 offers illustrative lines of inquiry for each of the potential evaluation objectives. These should be used to prompt project teams to consider the various issues that an evaluation can explore, though this list should not be considered exhaustive.

Lines of inquiry are a series of questions developed to provide greater direction on what one wants to find out – facts, opinions, experiences, unintended effects, etc.

When project teams generate their lines of inquiry there can be a tendency to list an endless number of questions. To aid in prioritizing this list, especially if it is long, consider how the resulting information will be used and by whom. Questions that do not inform an identifiable decision should be moved to the bottom of the list and, if resources are scarce, should be removed.

Illustrative lines of inquiry

| Evaluation Objective | Illustrative Lines of Inquiry |
|-------------------------------|---|
| Appropriateness Consideration | <ul style="list-style-type: none"> • What was the quality of the conflict analysis? • What is the link between the intervention strategy and the conflict analysis? Is it direct? • Does the strategy reflect key leverage points for change? • What is the theory of change? Was it articulated? • Was it appropriate vis-à-vis the context and the intervention strategy? • Are there other strategies/theories of change that could have contributed in a more significant manner? |
| Strategic Alignment | <ul style="list-style-type: none"> • How does the intervention contribute to the achievement of the organization's mission? • Does the intervention capitalize on the agency's unique expertise or experience? • Are there other organizations that could do this project better due to their expertise or situation? • Was the implementation reflective of the principles of the organization, for instance gender equality? |
| Management and Administration | <ul style="list-style-type: none"> • Was the direction, supervision and support provided to the intervention staff appropriate? • Were all aspects of the intervention well organized? • Were monitoring systems utilized to guide decisions and support reporting? • Were working relationships with partners effective (e.g., good communication, role clarity)? • Were all the activities run that were included in the project work plan? If not, why not? |
| Cost Accountability | <ul style="list-style-type: none"> • Were costs projected accurately and tracked regularly? • Were alternatives considered to maximize the use of funds when designing the project? • Did management decisions result in significantly higher costs? • Does the organization try to use economies of scale where possible? |

Illustrative lines of inquiry

| Evaluation Objective | Illustrative Lines of Inquiry |
|----------------------------------|---|
| Process Implementation Appraisal | <ul style="list-style-type: none"> • Did the development of the process benefit from the lessons from past experience? • Did the staff adhere to ‘good practice’ standards (where they exist)? • Was the process responsive to the changing context and needs of the stakeholders? • What lessons can be learned from this process? • Did the staff utilize monitoring mechanisms to inform their process? • How was the organizational approach or methodology incorporated into the work? |
| Output Identification | <ul style="list-style-type: none"> • What outputs were produced? • Were they of appropriate quality? • Were the outputs in line with the expectations based on the planning documents? If not, why? |
| Outcome Identification | <ul style="list-style-type: none"> • What intended outcomes occurred? • Did the outcomes align with the expectations based on the planning documents? If not, why? • What unintended positive and negative changes occurred ? • What part of the project was most important in catalyzing the change? • What was the process (environment, community, personal) that catalyzed the shift? • Was a prior smaller change required to happen first before this outcome could occur? |
| Impact Assessment | <ul style="list-style-type: none"> • Is there evidence of the outcomes of the intervention being utilized? • Is there evidence that a transfer of the change occurred from the participants to the wider community? • Were there any unintended negative or positive ramifications occurring due to this project? • What change in the conflict resulted due to the intervention? |
| Adaptability of Change | <ul style="list-style-type: none"> • To what degree occurred as a result of the intervention continue? • If the changes were not sustainable, why not? • What could have been done to sustain the results better? • When the conflict shifts to a new phase in the conflict life cycle, will these changes adapt and continue? • Did these changes affect participants’ perceptions of what is possible in transforming the conflict? |

AUDIENCE: Indicates who the primary audience (users) will be as well as the secondary audience (readers).

EVALUATION METHODS: This section includes the evaluation approach, scope and potentially all or some of the evaluation plan. It also references the existence of monitoring data and the organization's expectations or standards for evaluations.

The approach and the scope of the evaluation should be clearly described. If either requires the project staff to have a role in the evaluation, such as conducting interviews or shadowing the evaluation team, this should be stated. If an evaluation plan exists or there are preferences that will affect the development of the plan such as having a quantitative-qualitative blend of data collection tools, then this should be included. More detailed information on evaluation plans can be found on page 153.

This section should also make reference to the monitoring data that is available to the evaluators. What data has been collected, when, and by whom should be clearly indicated. Finally, if an organization has expectations or standards to which evaluators are to conform, these should be referenced. Expectations range from the requirement to include quotes and stories in the final report to process issues such as when the project team must be included in decision-making. If a formal document exists, attaching them as an appendix to the TOR is a good idea. (To learn more about strategies for avoiding pitfalls, see page 172.)

Expectations range from the requirement to include quotes and stories in the final report to process issues such as when the project team must be included in decision-making

Implementation Information

EVALUATION MANAGER: This details who will manage the evaluation. At a minimum, the person fulfilling this role will be in charge of hiring the evaluators, approving the evaluation plan, ensuring that milestones are met and signing off on the final evaluation.

LOCATION: This section is directly related to the evaluation scope. In effect, it provides the implementation plan of the scope. Indicating the location may be as simple as the physical sites of the project. With a multi-site program the staff team should consider if the evaluators need to access all sites or a representative sample. This is another section that can be finalized with the evaluators' input.

As a rule of thumb, the evaluation report takes about 30% of the total time allocated for the evaluation.

DELIVERABLES: This indicates the products that are required from an evaluation. While there are a variety of deliverables possible, the most common is a report. If a report is desired, the organization should outline the sections it expects to receive. A standard evaluation includes:

- Table of contents
- List of acronyms
- Executive summary of no more than five pages
- Overview of the context
- Organization and program background
- Evaluation methodology
- Evidence-based conclusions: These include the findings (e.g., data) and the analysis (e.g., what they think it means)
- Recommendations (if these were included as part of the evaluation objectives)
- Appendices
 - Appendix A – Terms of reference
 - Appendix B – List of those interviewed (if not confidential)
 - Appendix C – If there are dissenting opinions, they could be included in an appendix

As a rule of thumb, the evaluation report takes about 30% of the total time allocated for the evaluation. For projects operating on tight budgets, one way to stretch that budget is to shrink the expected deliverables. Options include:

- Limited report: A document that offers conclusions and recommendations only, omitting such areas as context and project history, methodology, appendices, etc.
- Presentations: Conclusions presented in the form of a presentation with brief handouts to cover the main conclusions
- Informal Format: Use bullet points, rather than narrative style

If a shorter document would increase readership internally, suggest a page limit to the evaluators. Alternatively, one can request a full report plus a summary document which generally averages ten pages in length and provides more detail than an executive summary. The latter option, although very useful since it provides content for a variety of audiences, requires more time from the evaluators and hence has budget implications.

In addition, the report language(s) and the format (electronic, bound hard-copy) in which it is to be submitted should also be specified. If the length of the report is an issue, be sure to indicate what that should be, such as no more than 50 pages for the body of the report, with a five-page executive summary.

Direct quotes or stories can provide insights on the conclusions and bring the generalizations alive. They can also be very useful for other materials such as brochures or the organizational website. If so desired, state specifically that these are expected to be included. This is often found in the Evaluation Expectations document. Note, however, that in conflict situations quoting people by name or using other identifying traits may put them at risk. Furthermore, good evaluation practice demands that the evaluators request permission to reference information disclosed in conversation. Be sure to discuss the desire for quotes with the evaluation team to ensure that this request is feasible within the conflict dynamic.

Finally, indicate who will be responsible for approving the draft report.

Finally, indicate who will be responsible for approving the draft report. This often includes the evaluation manager and project team leader, although such approval may require thematic technical assistance (i.e., review of specific sections or topics by specialists in those subject areas).

DURATION AND WORKING DAYS: The duration of an evaluation considers the period of time in which the evaluation is running, which is often a far longer period than the number of working days. For example, an evaluation which requires 20 working days to complete may start in May with the final report due in the middle of July. Thus, the duration is May to mid-July.

The terms of reference should detail the approximate number of working days needed to complete the evaluation. If international evaluators are being recruited, the working days should be broken down by the number of days they will spend in the country of the evaluation and the number of days they will spend doing work in their home country(ies). These decisions directly affect the budget since in-country days are more expensive. The number of working days depends upon the complexity of the evaluation, the types of information to be collected, the security situation, geography, and the size of the evaluation team.

Sometimes the evaluators need to be in-country to collect data during a specific time period. This could be due to key staff availability, the need to observe an event or project activity, or because of contextual factors such as elections or memorial days that should be avoided. If the in-country dates are fixed or, conversely, if there is flexibility in the schedule, indicate this in the TOR.

One way to estimate the number of working days needed is to allocate days to each task in the evaluation process. A generic task list can be found in the Evaluation Working Days Worksheet below. Since only some of these tasks will be included in every evaluation, this worksheet should be completed near the end of the evaluation planning process.

Evaluation Working Days Worksheet

| Tasks | Time Required |
|---|---------------|
| Discuss and finalize the terms of reference with project team and/or evaluation manager | |
| Review organization and project documentation | |
| Conduct literature review of trends and standards in the conflict transformation field | |
| Meet with project team to review evaluation process. Discuss the benefits and concerns related to the evaluation. | |
| Design evaluation plan and discuss with project team | |
| Design data collection tools | |
| Test data collection tools | |
| Travel to and from country | |
| Travel within country | |
| Translate data collection tools | |
| Meet with donors | |
| Data collection in site one | |
| Data collection in site two | |
| Joint evaluation team analysis | |
| Prepare for project team discussion on draft conclusions | |
| Draft conclusions presentation and discussion with project team | |
| Write draft report (rule of thumb for entire report: 30% of overall time) | |
| Edit and finalize report | |
| Conduct workshop on evaluation with headquarters staff and/or key project team staff | |
| Unexpected time delays (e.g., deterioration of the security situation that delays in-country travel) | |
| Total time required | |

DEADLINES: There are a series of deadlines to set throughout an evaluation including:

- Evaluator recruitment deadline: If recruiting externally, determine when the applications or proposals are due to the evaluation manager.

- Deadline for the proposed evaluation plan.
- The deadline for the draft report.
- The deadline for the final deliverables.

If the evaluation is complex, involves a large evaluation team, or is extended over a long period of time, having a larger number of process deadlines (milestones) is useful for the evaluation manager. These clearly communicate expectations and enable the manager to monitor progress against expectations.

It is smart to set the final report deadline a few weeks ahead of the real deadline since there is a tendency for evaluations to take longer than expected. When setting the deadline, consider proposal deadlines, strategic planning dates, donor reporting cycles and any other process that the evaluation information is expected to feed.

LOGISTICAL SUPPORT: Evaluators rarely operate totally independently of the organization they are evaluating. The organization can save precious time (and therefore money) by providing logistical support to the team. This support might include:

- Ensuring that the team receives key documents in a timely manner
- Organizing travel and accommodation
- Arranging meetings with the project team and key staff
- Identifying qualified translators or local staff to work with the evaluation team
- Providing identity cards where necessary
- Providing mobile phone(s)
- Providing administrative support such as photocopying, fax machines, and office space
- Helping to set up interviews or focus groups, if requested by the evaluation team. (See the discussion of pitfalls, page 172, for more information.)

When an organization does not have the spare capacity to provide logistical support, this should be clearly indicated in the terms of reference so that the evaluation team can plan accordingly.

The Evaluation Team

EVALUATOR'S ROLE: This indicates the role of the evaluation team. (See page 96 in the Evaluation Preparation chapter for further information.)

EVALUATOR'S RESPONSIBILITIES: This outlines the organization's expectations of the evaluation team during each stage in the evaluation process, from developing the evaluation plan to finalizing the report.

EVALUATOR'S QUALIFICATIONS: Many of the key decisions in this area are made in the evaluation preparation stage, such as whether to hire internal or external evaluators, whether the focus will be on local and/or international candidates, the size of the team, and the general types of experience required. The desired and required qualifications must be narrowed down for the terms of reference.

There are a few generic qualifications that should always be considered regardless of the evaluation objectives. These include:

- Evaluation expertise, in terms of both experience and credentials
- Experience working in conflict contexts
- Facilitation skills
- Oral and written communication skills
- Conflict transformation knowledge
- Country experience or, at a minimum, regional experience
- Language capabilities

In addition to the general qualifications that should be considered, there are three considerations that may impact the qualifications best suited for the job: evaluation objectives, gender balance and the conflict context. The following table, Evaluation Objectives & Illustrative Qualifications, offers examples of the different types of experience and knowledge one should consider in relation to each evaluation objective.

This table is not comprehensive nor should it be considered as a list of requirements. In reviewing the table below, the use of the terms “knowledge” and “experience” is intentional. The term “knowledge” implies a minimum understanding of the concepts, while the term “experience”

Evaluation Objectives & Illustrative Qualifications

| Evaluation Objectives | Examples of Qualifications |
|----------------------------------|---|
| Appropriateness Consideration | <ul style="list-style-type: none"> • Peacebuilding theories of change knowledge • Theory-based evaluation knowledge • Conflict analysis knowledge |
| Strategic Alignment | <ul style="list-style-type: none"> • Knowledge of the organization's history and principles • Knowledge of non-profit performance measurement • Facilitation skills • Conflict transformation knowledge |
| Management and Administration | <ul style="list-style-type: none"> • Operations experience • Knowledge of good management practice • Knowledge of non-profit performance measures |
| Cost Accountability | <ul style="list-style-type: none"> • Knowledge of good management practice • Basic understanding of accounting systems |
| Process Implementation Appraisal | <ul style="list-style-type: none"> • Conflict transformation knowledge • Tool (e.g., dialogue, problem solving workshop) experience • Observation experience |
| Output, Outcome, Identification | <ul style="list-style-type: none"> • Tool (e.g., dialogue, problem-solving workshop) knowledge |
| Impact Identification | <ul style="list-style-type: none"> • Research skills, ideally quantitative and longitudinal studies • Long term country knowledge |
| Adaptability of Change | <ul style="list-style-type: none"> • Country or regional knowledge • Conflict transformation knowledge |

It is often difficult to identify evaluators with the full set of required qualifications, particularly when one of the requirements is fluency in a language that is not widely spoken. There is often an inverse relationship between the length of the list of required evaluator qualifications and the length of the list of qualified evaluators. In other words, the more qualifications needed, the smaller the pool of appropriate candidates will be. When faced with this scenario, the organization should reflect carefully on its needs and may wish to reallocate qualifications between the required and desired.

Consider, for example, an evaluation of a peace journalism project in Burundi with a process implementation appraisal objective. Depending on the local situation, it may not require the evaluator to engage extensively with listeners in the community who mainly speak Kirundi. If the majority of the media products are produced in French, the evaluator would not need to speak Kirundi but could conduct the evaluation in French. Speaking the local language would therefore be desired but not required.

Think about the conflict setting when selecting evaluators. Some nationalities or identity groups may have more access or obtain more honest information from the participants of a project or local communities than others.

Also, consider gender balance on the team, particularly if the organization is committed to gender mainstreaming. Might different genders have differing access to information on the ground, either more limited or more open? Consider, for example, a reconciliation project between two villages in rural Kosovo. Due to the traditional cultural context of the area, male evaluators may not be able to talk alone with the women of the villages. This may cause the women to self-edit their comments in front of men or limit the data collection to the perspectives of the men of the villages only.

Finally, think about the conflict setting when selecting evaluators. Some nationalities or identity groups may have more access or obtain more honest information from the participants of a project or local communities than others. For instance, one would not want to have Palestinian evaluators assessing a project that involved Israeli settlers. In communities traumatized by violence, identity or nationality can be key factors in establishing trust and, therefore, access to useful information.

In addition, it is important to remember that the community or participants often see the evaluator of a project as connected to or part of the project being evaluated. As such, if the choice of evaluator implicitly suggests bias or allegiance to a group, this message may also be applied to the project or organization.

See on page 163, Frequently Asked Questions section for further information on selecting qualified evaluators.

Application Guidelines

This section details how interested external evaluators can apply for the evaluation. There are three common ways to recruit external evaluators: sole sourcing, short-lists and open tender. The first, sole sourcing, involves offering the consulting position to one person who is generally well known to the agency. This approach is time efficient, though it requires the organization to know a qualified candidate.

The second option is to create a short-list of candidates and to ask each of them to apply. This request should include the terms of reference, which detail the application expectations and procedures. The short-list can be developed through organizational networks, past consultants or by contacting other agencies for their recommendations.

This approach is less labor intensive than an open tender and more intensive than sole sourcing. It can be useful for those who do not have a candidate in mind and as a way of expanding the potential pool of evaluators for the organization. From the perspective of the short-listed candidates, taking the time to develop a proposal is appealing

if they know that there are a finite number of potential candidates. As such, it increases the chance of quality evaluators taking the time to develop a proposal.

The third approach is open tender, which is similar to the process used to recruit a staff member. The terms of reference are circulated widely and interested candidates are invited to apply. Casting the net wide can be a valuable strategy as it may unearth talent previously unknown. On the other hand, this approach can be time consuming if there is a high volume of interested candidates.

For both the short-list and the open tender approach, the application process can be as simple or complex as necessary. Consider what information you need to select the best candidate and design the process accordingly. The simplest and most time-effective route is to request a CV/resume. While this provides information on a candidate's work experience, it does not give information on the person's approach or aptitude, which is an important limitation.

Alternatively, one can request a short – two-page – proposal in addition to the CV/resume. The requested content of the proposal should illustrate the candidate's key skills and experience that are relevant to the evaluation. For instance, the candidate could contribute a draft evaluation plan or indicate the challenges they expect to face and how they would overcome them or they could be asked to articulate the principles that guide their work. Expectations should be kept fairly low if requesting an evaluation plan since it will only be a “best-guess” from the candidates, who are operating on the limited information available in the terms of reference.

Finally, one can request a complete proposal including evaluation plan, budgets, CV/resume and work plan. This can provide valuable information to the selection process, but it will require more time from the selection team. The rule of thumb is that the more one expects from the evaluators who bid on a project, the fewer applications will be received.

For both the short-list and open tender approach, the next step is to identify the two or three best candidates and contact them to assess their competency. For those recruiting from a local pool of candidates, requesting presentations on similar topics may also be an option. More information on assessing competency can be found in the Frequently Asked Questions section on page 163.

Budget Guidelines

It is good practice to include the evaluation's financial parameters in the Budget Guidelines section, since these can provide a sense of the project's size to prospective consultants which, in turn, can enable them

The rule of thumb is that the more one expects from the evaluators who bid on a project, the fewer applications will be received.

The project team, the evaluators or a combination can develop the evaluation plan.

to tailor their application accordingly. For larger budgets, one can include the actual cost figure of the evaluation or provide some guidance on the range of the budget: “The budget for this evaluation is between \$70,000 and \$85,000 Canadian dollars.” Generally, proposals will use the full budget provided. Consequently, if limiting costs is an important criterion in the selection process, it should be clearly indicated.

If the organization requests budgets to accompany applications, outline what should be included or excluded from the budget along with any other parameters.

Contact Information

If the evaluation manager or project team personnel are willing to field calls from interested individuals, include their contact information in the announcement. Taking inquiries from interested parties can help the evaluation manager gauge both the volume of interest in the project and the expertise available, and the evaluators can become better informed about the proposed evaluation. That said, this can take an inordinate amount of time and, if started, should be offered to all who inquire.

II. THE EVALUATION PLAN

What is an evaluation plan?

An evaluation plan provides a structured layout for designing an evaluation. It facilitates consideration of the key aspects of the evaluation and can be a useful communication tool among the various stakeholders. The evaluation plan also provides a reference point for the evaluation manager and project team to use to contribute to the evaluation design and to monitor the implementation.

Evaluation Plan

| | | | | | | | | |
|-----------------------|--|---------------------|---------------------------------|------------------------|-----------------------------|-------------------------|-------------------|-------------|
| Evaluation Objectives | Lines of Inquiry (Indicators, Standards) | Decisions to Inform | Means of Verification (methods) | Data Source & Quantity | Location of Data Collection | Conflict Considerations | Means of Analysis | Time (days) |
|-----------------------|--|---------------------|---------------------------------|------------------------|-----------------------------|-------------------------|-------------------|-------------|

To put it in context, an evaluation plan is to an evaluation as a work plan is to a project. The evaluation plan should be thought through as much as possible before getting started, yet it should have enough flexibility so that the evaluators can make alterations as they proceed and have the opportunity to identify better means or new opportunities.

Who develops the evaluation plan and when?

The project team, the evaluators or a combination can develop the evaluation plan. If it is to be developed by the project team, that fact should be included in the terms of reference. In such cases, it is advisable to still leave room for the evaluators to suggest improvements, particularly as they gain knowledge about the project and its context. If the project team feels strongly about certain aspects of the plan, such as the location of data collection, for example, be sure to indicate that to the evaluators.

Alternatively, the evaluation plan can be requested as part of the tendering process. The plan can only be as good as the amount of information that the evaluator has, so one should keep expectations reasonable. In this case, it should be expected that the evaluation team will wish to update it as more information becomes available. Finally, the plan may be developed from scratch once the evaluators are contracted.

Whether developed from scratch or as part of the tendering process, the plan should be created in conjunction with the project team whenever feasible. If there is room in the budget, plan development can be highly participatory and can produce a second deliverable of increased organizational capacity. If finances are limited, some strategically timed meetings should be held with the project team to gather their input and concerns before decisions are finalized. Engaging the project team so that they support the methods is a key component of having the evaluation conclusions accepted and applied in the future. See Evaluation Management page 174 for further discussion on this topic.

What does an evaluation plan contain?

While there are a number of different versions, the one depicted in this manual offers a structured approach that includes the core subject areas in most evaluation plans.

EVALUATION OBJECTIVES: These should come directly from the terms of reference. The evaluation team may raise useful questions that will help refine the objectives.

LINES OF INQUIRY, INDICATORS, AND STANDARDS: The evaluation objectives will dictate which of these three are needed. For instance, all results identification objectives must have indicators. If the indicators do not contain targets (e.g., “35% increase in freedom of movement” – with 35% being the target), standards of achievement must be developed for each indicator. Sometimes these

concepts are also blended. For example, lines of inquiry may be sufficient on their own or they may require standards and indicators within each inquiry area.

- **LINES OF INQUIRY:** Lines of inquiry provide the evaluators with greater direction about the information needed by the project team. Whether indicators or standards are used within each inquiry area depends on the evaluation objectives. See the table on page 142 for an illustrative listing of lines of inquiry.
- **INDICATORS:** Ideally, the indicators are drawn from the project design documents, though the evaluation team may wish to add to or improve them. Adding new indicators is fine as long as it is permitted by the grant agreement. Some grants require the implementing partner to gather information on specific indicators as part of a larger cross-agency monitoring system. Another consideration is whether or not the project monitoring system has been operating effectively. If so, data related to existing indicators will already have been collected, so by changing or deleting them, that information may no longer be useable.
- **STANDARDS:** There are two forms of standards, process standards and standards of achievement, and both are used as a comparison against which the evaluator can draw conclusions. Where the evaluation objectives include process implementation appraisal, management and administration, or cost accountability, process standards must be established. These standards may comprise techniques, steps, procedures, principles, or some combination of them all.

Peacebuilding practice does not yet have internationally accepted standards of practice against which quality may be assessed.

In fields like public health, there are international standards of practice that can be utilized. Peacebuilding practice does not yet have internationally accepted standards of practice against which quality may be assessed. Consequently, each organization needs to spend time reflecting on what constitutes high quality process to enable the evaluation to provide useful information. Experienced evaluators may be able to facilitate this discussion.

Standards of achievement are also used to compare findings against. When one of the evaluation objectives is results identification, the evaluator can draw on the targets within the indicator as a point of comparison. Consider this example: “An expected 35% increase in young Nepalese men from village Q who openly travel through the “other’s” community at night.” If the evaluator finds that there has been a 50% increase in such travel, the 15% difference is obviously far above the intended standard or target set by the program. The key point is that there is something to compare the evaluator’s findings against in order to draw a conclusion.

If there is no target in the indicator, time should be spent at this stage determining what would constitute success. Although largely driven by the project team, stakeholders' views are extremely valuable. The evaluator's experience can also be quite informative. Of course, purposely setting a low standard is not only unethical, it is also unlikely to work. An experienced evaluator will have had other scenarios to compare the current one with and will make note in the evaluation of the below-average target.

If there is no target in the indicator, time should be spent at this stage determining what would constitute success.

DECISIONS TO INFORM: In order to achieve the evaluation goal of improving peacebuilding programs practically and conceptually, the evaluation plan must clearly relate to the decisions or learning to be sought from the evaluation.

MEANS OF VERIFICATION (MOV): The MOV, or data collection method, is the way in which data will be collected. Different methods may be selected for each evaluation objective or line of inquiry/standard/indicator or one method may be appropriate to gather information against a number of them. Since each method must be developed and then tested, the more methods utilized the more time required for the evaluation. One can also use multiple methods to triangulate the information received as a way of verifying its accuracy.

The MOV, or data collection method, is the way in which data will be collected.

All social science data collection methods, such as interviews, questionnaires, document review, or focus groups are potential means of verification, as are participatory methods such as mapping or drawing. Further information may be found in the Methods chapter on page 204.

In the last several years, there has been increasing attention on developing conflict transformation-specific methods that will meet the special needs of this work. Many of these are peacebuilding-specific modifications of a social science methodology such as questionnaires, while others establish their own processes. As of 2005, none of these newly devised tools have taken precedence, nor have many been thoroughly tested. There is every indication that this area will continue to receive increasing amounts of attention, which will only improve and expand the options. Since this manual is focused on designing for results rather than on conducting the evaluations themselves, listing the various new tools is beyond its scope. Further information on peacebuilding tools may be found in the Methods chapter on page 204.

In the last several years, there has been increasing attention on developing conflict transformation-specific methods that will meet the special needs of this work.

The data source is where the information will be accessed

DATA SOURCE & QUANTITY: Closely tied to the MOV section, the data source is where the information will be accessed, whether from participants in the project, media professionals, judges or schoolteachers. Again, the data source can be different for each objective or line of inquiry/standard/indicator or these can overlap. The key question is how the evaluator will best access the information. This section should be as specific as possible. Indicating "women", for example, as the data source is too general and will hinder the subsequent decisions in the evaluation plan.

Is this realistic within the conflict context in terms of resources, opportunities and constraints?

This column should also give an estimate of the number of responses needed from that data source. For instance, 90-120 (quantity) questionnaires (method) from NGO staff members whose agencies currently work on the flashpoints in Belfast and Derry/Londonderry, Northern Ireland (data source). The quantity needed will depend upon the scope of the evaluation. If the evaluation is national in coverage or requires generalized conclusions, the quantity necessary will be higher.

LOCATION OF DATA COLLECTION: At a general level, where data is collected will already be detailed in the terms of reference. The parameters, however, are often based on areas larger than the evaluator may be able to cover, which means that a second set of decisions must be made to more precisely pinpoint where the data collection will take place.

CONFLICT CONSIDERATIONS: This section is intended to systematically insert the conflict into each layer of the evaluation planning. Fundamentally, one needs to ask throughout the development of the evaluation plan: Is this realistic within the conflict context in terms of resources, opportunities and constraints? The project team is particularly valuable in this regard since they not only know the answers but often are better equipped to identify the salient questions. As decisions are made within the evaluation plan, check them against these additional questions:

- What is the security situation? Will it restrict travel? Will it restrict access to people? Can some people move with more security than others?
- Can the evaluation team physically gain entry into the target community?
- Do special measures need to be taken in working with a community such as a highly traumatized group?
- What steps need to be taken to guarantee the safety of those involved in the evaluation?
- What are the prohibited or taboo questions?
- What are acceptable and unacceptable words? What language is considered politically biased?
- In which language will the evaluation be conducted? Languages can have political implications and should be considered carefully.
- What implications does the identity of the evaluator have within a community?

- Can the evaluation process include basic conflict transformation principles such as, “conflict is not always negative” or “generating a best alternative to a negotiated agreement (BATNA)” in order to support the goals of the intervention?
- What impact on conflict will the evaluation actions have?
- What physical infrastructure exists such as postal systems, telephone lines and accessibility of roads?
- What are the literacy levels of the target groups?
- Is this an appropriate method for the culture from which I am gathering data?
- What methods are considered valid in the culture of the evaluation audience?
- What are the cultural expressions (whether physical or verbal) of discomfort or unease?

MEANS OF ANALYSIS: How will the findings be analyzed? Consideration of the means of analysis is an important step yet it is one that is often forgotten when doing qualitative research. This is especially important to consider when the members of an evaluation team, who all bring specific skills and expertise to the effort, have different analytical approaches. Since the analysis approach selected affects the conclusions generated, a team member with a particular background may recommend an analysis method that works best for her/his substance area not knowing that use of that approach will omit important information from other areas. Particular thought should therefore be given to how differences of opinion about the analysis will be discovered and resolved.

TIME (DAYS): This column outlines the estimated time that will be needed to accomplish each task. It is a valuable reality check to ensure that the evaluation plan stays within the projected time period in the budget.

How do you develop the evaluation plan?

Creating an evaluation plan requires a blend of creativity and practicality – creativity in determining the best way to access the right information and practicality in terms of operating effectively within the constraints of time, finances, skills, and context. Generally, finding this balance will require some flexibility throughout the evaluation, but it does not eliminate the need for planning.

Note that the degree of rigor (exactness and complexity) within the research components of the evaluation plan is not that of a doctoral

student. Yet good research practice must be applied so that the conclusions are credible. One standard to apply is: Will this plan provide the organization with sound information that it can use to base decisions upon with confidence?

Evaluation Plan Example

The evaluation plan is for a project located in Bukavu, Democratic Republic of the Congo (DRC) that has been in existence for one year. The project consists of four radio programs whose goal is to help strengthen the people's commitment to the Intercongolese Dialogue. Two of the four activities have been selected for inclusion in the evaluation. The evaluation plan, developed by the evaluator, only shows one aspect of the evaluation and is intended to be illustrative rather than comprehensive.

PROJECT LOGIC:

Goal: Strengthen the people's commitment to the Intercongolese Dialogue

Objective One: Increase people's knowledge of the Intercongolese Dialogue

Objective Two: Increase people's understanding of the relevancy of the agreements to everyday life

Objective Three: Increase people's participation in the dialogue

Activities: Four radio programs: one soap opera and three talk shows of differing formats and themes

Definition: Intercongolese Dialogue means the agreements and dialogue process

One of the evaluation objectives in the terms of reference is output identification because the project team wants to ensure that the activities are producing deliverables in as intended. In addition, since the project is only a year old, outcomes may exist, but it also may be too early to identify strong evidence of them. Consequently, collecting solid output information is helpful in terms of ensuring that the team is building toward what they hope to accomplish.

The line of inquiry associated with the objective is, "What outputs have been produced by this program over the past year?" More lines of inquiry could be included if the project team needed other information to help them make decisions. For instance, a line of inquiry that looked at the quality of the outputs could be added.

Will this plan provide the organization with sound information that it can use to base decisions upon with confidence?

Evaluation Plan Example

| Evaluation Plan | | | | | | | |
|------------------------------|--|--|--|-----------------------------|--|--|-----------|
| Evaluation Objectives | Lines of Inquiry (Indicators, Standards) | MOV (methods) | Data Source & Quantity | Location of Data Collection | Conflict Considerations | Means of Analysis | Time |
| Output Identification | <i>Line of inquiry:</i> What outputs have been produced from this program over the past year? | | | | | | |
| | 80% of media staff able to independently produce a radio show | In-person interview | Head of radio production Country director | Bukavu NGO office | | Review of interview notes | .25 day |
| | 280 minutes/month of radio program Y aired during prime time for the target audience | Document review Radio listening | Weekly production log (one/month) One station/radio program | Bukavu NGO office | | Collate the consistency of broadcast times and radio minutes Verify broadcasting by tuning in at the right time for each show | .5 day |
| | 50% of radio station listeners repeatedly listen to radio program | In-person interviews | Radio manager(s) from each of the 5 stations | Throughout Bukavu city | Female evaluator should not travel alone Is the population settled enough for these numbers to be gathered? | | 1.25 days |

The indicators were found in the program design tool which was a logical framework. In this case, the indicators apply to both activities included in this evaluation. The original indicators were:

- Number of individuals able to independently produce a radio show
- Number of minutes of radio programs aired/month during prime time for the target audience
- Number of listeners/radio program

If the activities were different, such as a youth camp and radio programming, the output indicators would need to be unique to each of those activities. Minor changes were made to the first two indicators to make them more specific and to avoid any misinterpretation. The third had to be altered to meet the realities on the ground, although its essence was kept.

None of the indicators contained targets, so the evaluation team established the standard of achievement at this stage. This process included individual discussions with the project team, emails sent to headquarters support staff, former staff and to the regional media manager. They were asked to reflect back on when the project was designed and their expectations at that point. Since it is difficult to separate what one knows now from what was known when the project began, all of these responses were brought together and blended with the expertise on the evaluation team. The results were then brought back to the current project team to ensure that they accepted them as an appropriate standard.

The data on the first indicator, “80% of media staff able to produce a radio show,” would be found through two in-person interviews (method), one with the head of radio production and one with the country director (data source and quantity). Since this is a new office, open only for one year, the team is small enough for these the head of radio production and the country director to work directly with each staff person. Therefore, the evaluators felt confident that they were an adequate source of information; however, it would be necessary to verify this in order to rule out any potential staff politics at play. To this end, the evaluators reviewed the radio production log to determine the number of shows produced by each staff person, frequency of production, and variety of working partnerships. The latter was to ensure that a team had not formed where one person actually produced the show without much assistance from the other. Note that, if there was a quality component to this indicator (e.g., number of individuals able to independently produce high-quality radio shows), standards would need to be established to indicate what constitutes “high-quality,” and the evaluators would then need to assess the staff against these standards.

The two interviews can be conducted in the Bukavu office (location of data collection), though they should be held behind closed doors since they may involve anonymous comments about staff competency. The means of analysis in this case is very simple and can be a comparison of the interview notes. This will answer the question, “Do we need to provide additional staff training in radio production?”

The evaluation team questioned the second indicator, “280 minutes/month of radio program ‘Y’ aired/month during primetime for the target audience,” in terms of whether it was a result of or simply a part of project activities. After discussion with the project team, it was accepted as one of the indicators to be used because the project produces shows and does not have its own radio station. As a result, part of the achievement in this case is convincing radio stations to air their broadcasts during the target audience’s prime listening time. Of course, if the organization pays the station to play its programming, this would not be an indicator since would simply be part of the activities.

The data on this indicator can be gathered by reviewing (method) twelve weekly production logs (data source and quantity) – one per month. One does not need to review every week, unless it is evident from the sample selected that there is great variance in the information. Since the monitoring needs informed the creation of the production logs, data on when the radio show was broadcast each week is recorded in the log. To verify that this is accurate, the evaluator could tune in (method) to each of the two programs (data source and quantity) at their scheduled broadcast times. This can all be accomplished from the Bukavu office of the NGO (location of data collection). The production log information would be collated to determine the consistency of the broadcast times and the number of minutes added up for an approximate total (means of analysis).

This information will indicate to the project team whether or not they need to do additional work with the radio stations in order to get the correct time slot for the target audience. It will also contribute valuable information to the NGO’s understanding of the intensity and frequency of messages required to initiate change.

The third indicator is “50% of radio station listeners repeatedly listen to radio program Y.” Although seemingly straightforward, it is very difficult to gather accurate data for this indicator. If the evaluation budget and time were sufficient, one could conceivably gather this data in a normal context. However, Bukavu has limited road infrastructure, a highly insecure environment outside of the city itself, and an extremely mobile population (conflict considerations), all of which would make it difficult for the team to generate reliable conclusions. With sufficient time and money, these challenges could be surmounted to some degree; however, this budget did not offer that opportunity.

The fall-back alternative is to conduct interviews (method) with the radio managers (data source) at all five stations (quantity) to seek their assessments of the percentage of listener market share each of the stations has. Since there are no land lines in Bukavu and few cell phones, these need to be in-person interviews at the stations (location). In addition, the approach to these interviews should be flexible because there may be multiple managers or other knowledgeable personnel at each station, which may require small group interviews to be conducted instead. Due to the high levels of violence against women, the female evaluator should not travel alone to these stations (conflict consideration).

Although it is almost a certainty that the stations will not have statistically generated ratings, they will have an informal sense of their stations' audience and the popularity of each show. This data should be supplemented by including questions in the other aspects of the evaluation to triangulate the estimates. The means of analysis will be to compare the radio station estimates of listenership with the other data gathered. The evaluation team should indicate these trade-offs to the project team so that they understand the limitations of the information. All of this information will feed into the organization's decision regarding whether or not to continue airing each of the radio programs.

III. FREQUENTLY ASKED QUESTIONS ON WORKING WITH EXTERNAL EVALUATORS

Once the decision has been made to use an external evaluator (see page 126 for more information on this decision), a variety of questions are commonly raised. This section provides some tips and tactics for answering these questions.

1. When should I start recruiting to get a quality evaluator?
2. Where do I find evaluators?
3. How do I appraise the competence of external evaluators?
4. Is there anything different in an evaluator's contract?
5. What do I need to do before the evaluation begins?
6. How do I know if the evaluation plan is a good one?
7. The security situation has worsened and it is not safe for the evaluation team to travel to some of the worksites. Should we cancel the evaluation?
8. Who should manage the evaluators?
9. How much management is needed for external evaluators?
10. Can I participate in the evaluation of my project?

11. What should I do if I don't agree with the draft report?
12. What do we do with the evaluator's primary data, such as interview notes, after the evaluation is completed?

1) When should I start recruiting to get a quality evaluator?

Whichever approach is selected for recruitment, sole-source, short-list or open tender, professional evaluators with track records of quality performance generally have booked schedules. Therefore, to obtain an experienced professional, it is recommended that the recruitment take place as far in advance of the evaluation as possible. In the ideal scenario, development of the TOR should begin six months before the time the team is expected to arrive on the ground. As a rule of thumb, there should be two months between finalizing the contract with the evaluator and the on-the-ground dates of the evaluation.

Consider, as an example, an evaluation that is to occur in June. The TOR would be developed in January/February, the recruitment process would occur in March, and the contract finalized no later than early April. In May, the preparation should start so that the team can be on the ground in June. Shorter timelines are possible, but the organization will pay the cost in terms of both the intensity of work required to plan the evaluation, which must be squeezed in among regular job responsibilities, and a likely decrease in the quality of evaluators available.

...professional evaluators with track records of quality performance generally have booked schedules.

2) Where do I find evaluators?

There are a number of ways to connect with prospective evaluators including evaluation associations, M&E websites, universities or conflict transformation networks.

Many countries have national evaluation associations that offer listservs, job postings on their websites, or on-line databases to connect with their membership. There are also a few M&E websites that post opportunities for evaluators. Check Appendix B on page 228 for a listing of these websites and a selection of national evaluation associations.

To facilitate the sole-source and short list approaches, it is recommended that organizations maintain a database of the evaluators it has used. Include feedback from the project team and evaluation manager on the professionalism, quality and performance of the evaluator to inform future decisions. Names of individuals who have not performed to a reasonable standard should also be recorded to prevent the organization from making the same hiring mistake in the future.

3) How do I appraise the competence of external evaluators?

There are many ways to appraise an evaluator's competence. Some of the more common steps are included here. The first step is to compare the person's application against the qualifications listed in the terms of reference. Be sure that she/he has substantive and solid experience in the required areas as evidenced by the CV/resume. A common mistake is to rely on the cover letter alone, which may provide a different picture from the CV/resume itself.

The next step is usually an interview with the evaluator or the evaluation team. In this discussion, seek to determine the depth of the person's expertise on evaluation as well as her/his core substantive areas of expertise (such as child soldiers or ADR processes). Find out if the evaluation team wrote the proposal or application; if not, be sure that the proposed evaluators are as qualified as the application writer. In addition, keep in mind that the existence of many completed evaluation assignments on a CV/resume may not be an indicator of quality. As such, it is a good idea to request a sample evaluation to review, though this may be more difficult than expected because the majority of evaluations are the property of the organization evaluated. When reviewing a previous evaluation, check to see if it has the elements of quality that the organization has determined it needs for its proposed evaluation.

When using an open tender or short-list approach, it is always a useful final step to check references. In these conversations, be sure to ask about:

- **Timeliness:** Did they meet the milestones and deliver the product on time?
- **Responsiveness:** Were they flexible enough to deal with unexpected challenges or delays? Did they respond to the project team in an appropriate manner?
- **Relevance:** Did they follow the terms of reference?
- **Professionalism:** What was their work style, communication ability, or degree of cultural sensitivity?
- **Evaluator Good Practice:** Did they engage openly with the project team? Did they proactively explain their decisions based on good practice in evaluation?
- **Evidence-based Conclusions:** Were the report's conclusions evidence-based or mainly conjecture?

A common mistake is to rely on the cover letter alone, which may provide a different picture from the CV/resume itself.

- **Conflict Sensitivity:** Did they display a sound understanding of the implications of doing evaluations in conflict areas?

A common mistake is to simply ask the referee if she/he was happy with the evaluation. This assumes both a level of knowledge about good practice in evaluation and that the two organizations have similar expectations of quality. In some cases, the referee may be happy with the product simply because it reflected well on her/him rather than because the quality of the evaluation was high.

4) Is there anything different in an evaluator's contract?

The majority of a contract for external evaluators is the same as for any consultant. Remember to include the terms of reference as an appendix and refer to it in the body of the contract. If the organization has developed standards or expectations for what constitutes a quality evaluation, these should also be included either within the contract itself or as an appendix.

A section should also be included that indicates how the evaluator can use the information after the project is completed. It is most often referred to as "confidentiality" and it should state precisely who owns the data and the final report as well as how these may be used in the future. For instance, if the organization owns the final report from an evaluation, the evaluator would need to request permission to use it as a sample in an application for another assignment. The same holds true for posting a report or information derived from the evaluation on a website, quoting it in other reports or making public statements about the conclusions.

A common mistake is to simply ask the referee if she/he was happy with the evaluation.

5) What do I need to do before the evaluation begins?

A number of actions must occur between the time that the evaluation team is hired and the start of the evaluation. The first step is to gather all relevant documentation for the evaluation team. This almost always includes the funding proposal(s), donor reports, baseline and/or previous evaluations, program background information, organizational history and monitoring data. The challenge is to provide critical documentation that will inform the team but which will not overwhelm them. Be sure to create a listing of key staff, titles and contact details within the project if one does not already exist.

The evaluation will proceed much more smoothly if it has been discussed with the project staff before the evaluator arrives in their office. Transparency and communication around the motivations for the evaluation, the role of the evaluator, and the potential uses of the final report can allay

Transparency and communication around the motivations for the evaluation, the role of the evaluator, and the potential uses of the final report can allay natural fears that may arise.

natural fears that may arise. These discussions should encourage staff to be open and honest with the evaluator. In this process, individuals who oppose the evaluation may appear, which may require that individual or more focused discussions be had with them prior to the evaluation to allay their fears.

It may be useful to think about which key stakeholders the evaluators should meet beyond the project staff. The list of these stakeholders should then be reviewed with a conflict lens. For instance, will involvement in the evaluation compromise the position of any of these people? Could their participation place them in danger in their community? Are any of the stakeholders so badly traumatized that the evaluators should be careful not to negatively affect these individuals? Is there anyone who may not be willing to meet with strangers (i.e., the evaluators) or trust them enough to speak honestly and would therefore need a personal introduction from a project staff person? Having this list available with any conflict considerations will be very helpful to the evaluators.

Finally, pending what logistical support was promised in the terms of reference some logistics may need to be addressed such as the purchasing of airplane tickets.

6) How do I know if the evaluation plan is a good one?

So, you have left the methodology open for the evaluators to construct and they have submitted an evaluation plan for approval. Here are some of the questions to ask:

- Has each of the evaluation objectives been included?
- Are the lines of inquiry supplemented where needed by standards or indicators?
- If new indicators were added, how are they better? Are they “good” ones?
- Are all the necessary locations covered?
- Are the research methods appropriate for the context? For instance, is a questionnaire suggested for a location that has a low literacy rate? Are focus groups planned in areas where women cannot attend?
- Do the research methods appear to be the best ones for obtaining the data? For example, focus groups are not appropriate if community members tend not to speak openly and honestly in public.

- Is there a quantitative-qualitative blend in the research methods?
- Do the methods triangulate?
- Are there any concerns with accessibility to the data source groups?
- Does the amount of time allotted for the evaluation seem feasible? Are there more days planned than are available for the evaluation? Has the time required for local travel been taken into account? Does the plan account for the extra time that translation requires?
- Is existing monitoring data being used?
- Have they ensured that different sources or different processes have verified the monitoring data?
- Are the conflict considerations accurate, necessary and comprehensive?
- Are the means of analysis benefiting from the full team's expertise?

If the organization has M&E technical assistance available from staff members who are not already involved in the evaluation, this is certainly the time to request their input. If there is no in-house expertise, consider hiring an M&E advisor who can be consulted at key steps throughout the evaluation process. To save time, the same advisor can be used for all evaluations.

7) The security situation has worsened and it is not safe for the evaluation team to travel to some of our worksites. Should we cancel the evaluation?

The answer to this question is, "It depends." When the safety of the team is at stake, one should always consider postponing to a later date. Under circumstances of questionable safety, there are at least five other options to consider: changing the methods, excluding locations, reframing who travels, considering proxies and using secondary sources.

First, are there alternatives to the planned data collection methods that do not require travel to the dangerous area? Consider using phone interviews or a staff person who is already in the area as ways to conduct the conversation and provide notes to the evaluation team. If a case study is being written, key actors can be asked to chronicle their experiences if this would not put them in danger.

When the safety of the team is at stake, one should always consider postponing to a later date.

Second, can other worksites be used without compromising the results to a great degree? For instance, a nationwide survey on media attitudes in Burundi excluded from the sample the three provinces that were inaccessible because of poor security conditions. Though such exclusion might not be the ideal, it is still a credible option as long as the limitations in the data collection process are clearly communicated in the final report and the conclusions qualified accordingly.

Third, can the people who are to be interviewed travel safely to the evaluation team? In certain cases, movement is possible for one community but not the other.

Fourth, are there proxies available who can speak to the evaluator(s) on behalf of the target population? Talking to parents of child soldiers who are fighting in the bush for their perception on the child's attitudes and experience is an example. Clearly, this is not a perfect strategy; however, if one has no way of accessing child soldiers, it may provide insights and information that, when triangulated with other information, becomes useful in decision making.

Fifth, shift the data collection to existing sources of information, such as other NGO studies or internal reports, academic research papers, evaluations for other projects or other organizations, newspaper articles, etc. The exclusive use of secondary sources is a blunt instrument, but if enough of it can be collected, a useful picture can emerge on issues addressed by the evaluation. Which of these strategies makes sense will depend on the evaluation objectives and intended use.

8) Who should manage the evaluators?

Appointing an evaluation manager who is not on the project team of the project to be evaluated is the most effective strategy. When done well, the separation of evaluation management from the actual evaluation creates a “political firewall” between the evaluators and the program, which keeps the evaluation free from the control or undue influence of those responsible for the project. Moreover, this structure enables an internally driven evaluation to be as close to an independent evaluation as possible.

Appointing an evaluation manager who is not on the project team of the project to be evaluated is the most effective strategy.

The evaluation manager is appointed at the beginning and is named in the contract as the responsible party. She/he works with the program team on all key evaluation preparation decisions and on implementation and use (stages one, two, and three). The person in this position facilitates development of the terms of reference, runs the recruitment process, manages feedback on the draft report, signs off on the completion of the final report and authorizes final payment. Good communication by the manager with the program team is key to producing a useful document.

The manager needs to keep squarely in mind that she/he is managing the administration and not the substantive content. The nature of the conclusions is outside the remit of the manager, unless the conclusion does not seem evidence-based or is missing key variables. In that case, it is appropriate for the manager to request an alteration to the report in the form of either additional evidence to support the conclusion or the incorporation of the missing variables, provided that it had been verified.

In a mixed evaluation team, with internal staff and an external consultant, it is helpful if the manager is not the internal staff representative. Separation of those roles will decrease role confusion and possible tension. Where management of an evaluation becomes more difficult is when the individual managing the evaluation is also part of the project being evaluated (see the ethics chapter page 188).

9) How much management is needed for external evaluators?

External evaluators should be managed in much the same way as any external consultant under contract. Ensuring that the team is using its resources wisely, staying aligned with the terms of reference and meeting deadlines should all be managed by the organization.

The management role is a very active one from the preparation stage through the data collection stage. Prior to the point when evaluators arrive on site, the manager should carefully review the draft evaluation plan. Although communication will be ongoing while the team is conducting the evaluation, the manager should hold a final discussion with the evaluators before they leave the field location to ensure that all of the necessary data has been collected. This discussion should also identify for the project team what the next steps will be, including the submission of draft conclusions. For a complicated or lengthy evaluation, periodic milestones should be agreed upon and the manager should regularly check progress against these milestones. The manager continues to ensure that deadlines are met and report specifications are maintained after this point, but the nature of the report's conclusions is outside of the management mandate.

The manager needs to keep squarely in mind that she/he is managing the administration and not the substantive content.

10) Can I participate in the evaluation of my project?

This decision should be made in the evaluation preparation (stage one) and all project staff members should be advised about their respective roles. If it is a self-evaluation, a participatory evaluation, or you are a member of a mixed evaluation team, then you will likely be very involved in the evaluation process. Alternatively, if the evaluation has

been structured to include a capacity strengthening element (similar to a participatory evaluation), there will be structured ways for you to participate.

As a member of the project team, you may also be invited to meetings with the evaluation team. This would keep you informed of the progress and decisions being made. Finally, the evaluation team should be required to provide feedback to the staff before they leave the field. This feedback can range from initial impressions to sound analytical conclusions, depending on whether the in-country team has had time to do the analysis. This is an important step since it provides the staff with an opportunity to question, understand, and clarify different conclusions.

Beyond these relatively structured forms of participation, inclusion in the evaluation process should be left to the discretion of the evaluation manager, who will consult with the evaluation team. Direct requests to participate from the project staff to the evaluators can be awkward and perceived to be power-politics at work. Be respectful of the evaluator's role and the need for the conclusions to be sound and the evaluation credible.

11) What should I do if I don't agree with the draft report?

The principle that cannot be violated with an evaluation is that the conclusions must be based on the evidence. These are not the views and opinions of the project staff that have been foisted onto the evaluation team. Nor should evaluators be drawing conclusions that are not based on the data.

That said, there will be instances where the project team does not agree with the evaluation conclusions. If there is disagreement because the conclusions are not supported by evidence, a request can be made that the evidence be included or the conclusion omitted. Of course, this assumes that the evaluation terms of reference included evidence-based results. If it appears that the evaluator was not aware of critical information when she/he analyzed the information, it is appropriate to inform the evaluation team of these extra variables and how you believe it impacts the conclusions.

If the disagreement is with the conclusion itself and the evaluator has heard the concerns and decided not to alter the conclusion, one alternative is to include the organization's view as part of the evaluation document. This can be done in the text, as a footnote, or as an appendix to create a document that shows both perspectives. Including the differing views is extremely important because the next evaluation will likely ask what happened to the conclusions in the previous evaluation.

Direct requests to participate from the project staff to the evaluators can be awkward and perceived to be power-politics at work.

The principle that cannot be violated with an evaluation is that the conclusions must be based on the evidence.

Editorial control of the final report is often indicated in the terms of reference. Generally speaking, the last word on the final evaluation report should be left to the evaluators. If another version of editorial control is selected – though this is not recommended – it should be specified in the terms of reference.

12) What do we do with the evaluator’s primary data, such as interview notes, after the evaluation is completed?

Primary data, such as interview notes, should be destroyed if the evaluator promised confidentiality to the sources in the process. This is a particular necessity in situations of conflict where the opinions and stories collected may be about sensitive issues and, at times, even illegal ones.

Hard copies of surveys and questionnaires should be stored with the analysis and conclusions so that future evaluators can use the raw data to verify the conclusions of the first evaluation or to contribute to a different study to save resources. If the data has been entered electronically into a software program, this too should be saved, again to eliminate the need to re-enter it for future uses. If the surveys or questionnaires include the responder’s name, the organization must take extra precautions to store that information in a manner that protects the responder.

Generally speaking, the last word on the final evaluation report should be left to the evaluators.

IV. STRATEGIES FOR OVERCOMING COMMON EVALUATION PITFALLS

There are many ways in which evaluations can falter along the way that can hinder the quality, usefulness or application of the evaluation conclusions. With some preplanning and creative thinking, these common pitfalls can be avoided.

“When the eyes are bigger than the plate”

It is common for evaluation managers and project teams to design an evaluation that is far too large for the proposed budget. Creating too many evaluation objectives or having too broad a scope for what is feasible are the most common ways for this to happen. To avoid this problem, be sure to work through the key evaluation preparation decisions

in the project design process, as outlined in the Evaluation Preparation Decision flowchart on page 99. This will enable the project team to create a sound budget estimate as part of the project proposal.

If the evaluation appears to be too large for the budget and the team is having difficulty scaling back, draft the terms of reference in a flexible way until the evaluators have been contracted. Then, during the first meeting between the evaluators and the project team, they can work collaboratively to finalize the terms of reference with an eye towards feasibility.

“Don’t shoot the messenger”

One of the many benefits of evaluation is that it helps to identify issues and information for the project team that were previously unknown. In this process, both positive and negative points can be unearthed. Sometimes the negative findings can be quite challenging and even threatening for a project team to handle and there can be a tendency to resist the resulting conclusions and recommendations.

There are two common responses to negative conclusions from an evaluation: discount the methods or criticize the evaluator. In the first of these, the validity of the methods used in the evaluation will be questioned. Statements like, “Of course that is what you found; you talked to the wrong people!” or “I’m not surprised they didn’t tell you otherwise since you spoke to them in a group and they would never contradict the norm in that setting,” depict this well. A strategy that may help prevent such reactions is to engage the project team in the methods discussion. This discussion can be as participatory as is feasible and it should be made clear that this is the appropriate time to challenge the methods. At a minimum, a discussion based on the evaluation plan should be held with the full project team with the explicit purpose of getting the team to reach consensus on the best strategy.

The second common response to negative conclusions is to challenge the evaluators’ credibility. Statements like, “Well, this field is not their specialty, so they just don’t get it” or “I didn’t like him from the beginning,” are often used to undermine the evaluators and, therefore, their conclusions. The best way to stop this before it happens is to hire evaluators who behave in a manner that instills confidence in, and respect for their professional competencies. If there is a concern that there may be resistance to doing the evaluation among the project staff, this should be communicated up front to the evaluators and their strategies for dealing with such resistance should be discussed. For instance, in certain cultures it would be important for the evaluators to articulate their credentials to establish credibility with the project team. The evaluation manager can play a key role in this through regular check-ins with the evaluators and the project team.

There are two common responses to negative conclusions from an evaluation: discount the methods or criticize the evaluator.

“Whatever you say, ma’am”

It is important for an evaluation manager to be conscious of the politics of evaluation throughout the process. Recognize that evaluators are often consultants whose living is dependent on client satisfaction and referrals. As such, there are instances where pleasing the client may be put ahead of good practice. This is most commonly seen in evaluation reports that omit or play down problems that were identified, but it may also occur in the selection of methods – for instance, using a questionnaire because the client thinks it is the only credible data collection tool when qualitative information is actually what is sought.

To avoid this, evaluation managers should be sure to ask about professionalism during the reference check. It also helps to communicate openly with the evaluators that the organization seeks to learn not only about what they do well, but also about the areas in which they need to improve. An evaluation report that only addresses the positive but not the negative (or vice versa) will not be deemed satisfactory. At the same time, the evaluation manager should communicate the same message to the project team.

An evaluation report that only addresses the positive but not the negative (or vice versa) will not be deemed satisfactory.

“Being all things to all people often means being nothing to anyone”

There are many different stakeholders in peacebuilding projects – participants, project staff, the organization, partners, and donors – all of whom have different information needs when it comes to an evaluation. A common pitfall is to identify all these groups as the evaluation audience. While a donor may want to know if the organization being evaluated met the donor agency’s mandate, the project participants may be most interested in sustainability or unintended negative effects. The project team on the other hand may want feedback on the process utilized. When making the key evaluation decisions at the project design stage, it is important to think through the audience question. Determining who will be the “users” versus the “readers” of the evaluation is effective in identifying the primary audience. See page 108 in the Evaluation Preparation chapter for more information.

“We’d be happy to do that for you”

Project staff members commonly draw up interviewee lists and offer to set up the interviews for an evaluation. This can appear to the evaluators to be a great way to save them from the tedious job of arranging appointments. However, project teams may select people who they think will speak most positively about their work or, when arranging the interview, they may intentionally or unintentionally tell the person what to say to the evaluator.

Conflict situations are highly volatile and dynamic; therefore, programming often needs to shift or, at times, change focus entirely in order to remain relevant to the needs on the ground.

Consider, for instance, the Sudan IDP-host community example used in the Evaluation Preparation chapter on page 109. Bias or favoritism might become a factor if the project team handpicked those individuals whom they felt had gotten the most benefit from the project and in arranging an interview said, “Be sure to tell the evaluators how important this was to you and how much this changed your life. We really hope to get more funding and we only will receive it if the evaluation is good.” Before the evaluation even arrives, the individuals have been told what to say.

In some cases, having project staff do these tasks may be necessary if the budget and/or time is limited. In these situations, the evaluator should identify the generic types of people to interview and then provide the project staff with a script or guidelines for their discussions with the potential interviewees. In addition, the evaluator should verify that there is a wide spectrum of people on the interview list and allot extra time for spontaneous conversations when she/he is actually on the ground.

“We want Wonderwoman/Superman”

For complicated or multi-faceted evaluations, it may be difficult to find one person who possesses all the skills and experience needed. One strategy for addressing this difficulty is to devise a team on which all the required skills are distributed among the different team members. To ensure the effectiveness of this strategy, it is critical that the team structure itself and the evaluation process to capitalize on the diversity of expertise among the members. There are four options within this strategy:

- Having daily debriefs to do real-time analysis of information
- Jointly constructing the questions in the research instruments to ensure they include all key perspectives
- Identifying key issues from the various perspectives represented on the team during the analysis phase
- Joint report writing

“That logframe is so out of date”

Conflict situations are highly volatile and dynamic; therefore, programming often needs to shift or, at times, change focus entirely in order to remain relevant to the needs on the ground. Very often the program content is modified while the original logical framework is left untouched. Changing the content without changing the logical framework can create a dilemma for the evaluators because many evaluation design decisions

are based on the logical framework. It can create even greater problems if the evaluation objectives include results identification or adaptability of change since the results stated in the logical framework will no longer be relevant to the new work on the ground.

The easiest way to avoid an outdated design tool is to use them as they are intended to be used. As such, when the conflict situation shifts, use the design tool to assist in the creation of the new focus. If it is too late and the evaluation is about to start, the development of a logical framework that is representative of the new programming can be included in the terms of reference. This additional task would then need to be factored into the budget, timing and skill set of the evaluation team, yet it can contribute another useful deliverable to the program in addition to the evaluation. It is recommended that the program team be involved to some degree in the creation of the logframe to ensure its accuracy and utilization by the team beyond the evaluation.

“It’s not a result unless it’s a number”

Traditionally, results have been considered credible if they were generated through quantitative research methods. This approach can only tell part of the story and it leaves many questions unanswered. For example, a questionnaire consisting of multiple choice questions may provide the project with information on how many people attended a town hall meeting and how much they learned from it, but it would not be useful for conveying any of the unique circumstances that brought them there or that caused them fear or concern in attending. This quantitative focus is partially due to the cross-over from the development field where evaluations of infrastructure or public health projects were handled almost exclusively in quantitative form.

Because conflict transformation projects often deal more in the realm of qualitative changes, such as decreased feelings of hatred or changes in attitude toward the “other,” a quantitative-qualitative blend is best suited to grasp the difference the program is making in its entirety. This expectation should be communicated in the terms of reference and should be clearly visible in the evaluation plan.

“I’ll just be a fly on the wall”

Some project staff may be very interested in the evaluation process and/or in what people have to say about the intervention and, consequently, may request or even insist on shadowing the evaluation team. Pending the evaluation objectives, identified use and team makeup, this interest may not be a problem and, if planned for, could be a capacity-strengthening exercise. However, in other cases it can cause difficulties.

Because conflict transformation projects often deal more in the realm of qualitative changes, such as decreased feelings of hatred or changes in attitude toward the “other,” a quantitative-qualitative blend is best suited to grasp the difference the program is making in its entirety.

Staff members who sit in on interviews or take part in the evaluator’s debriefs can hinder the openness of these discussions. In some instances, staff presence can change the entire dynamic of an interview, to the point that the interviewee alters her/his story to ensure ongoing NGO support. Moreover, if the audience or use of the evaluation requires that there be a high degree of credibility in the methodology, permitting staff attendance in certain aspects of the evaluation may reduce overall confidence in the evaluation.

The easiest way around this pitfall is to include in the terms of reference development a discussion on the role of staff. That role must then be communicated clearly to all project staff. It should be noted that, if staff insist on accompanying the evaluators, a professional evaluator should immediately communicate the potential pitfalls to the evaluation manager.

“But I thought we were on the same page”

Since the fields of evaluation and conflict transformation are still figuring out exactly how they best fit together, a set of commonly held norms or best practices for evaluation have yet to be established. As a result, any assumption that the evaluator and the organization are on the same page regarding what constitutes “good practice” for the evaluation is a pitfall waiting to happen. One cannot assume, for example, that meeting with the project team to discuss the terms of reference or doing a debrief before leaving the country will take place unless it is specifically stated in writing.

To avoid disappointment, organizations should be explicit about their expectations of the evaluation process. It is best to document these expectations and include them as part of the contract. If that is not possible, then at a minimum, have a conversation with the evaluator prior to each stage of the evaluation process about what you as the client consider to be the norm. Some of the areas to consider are the alignment with organizational language, key steps in the process, ethics standards, report content, and expectations and confidentiality.

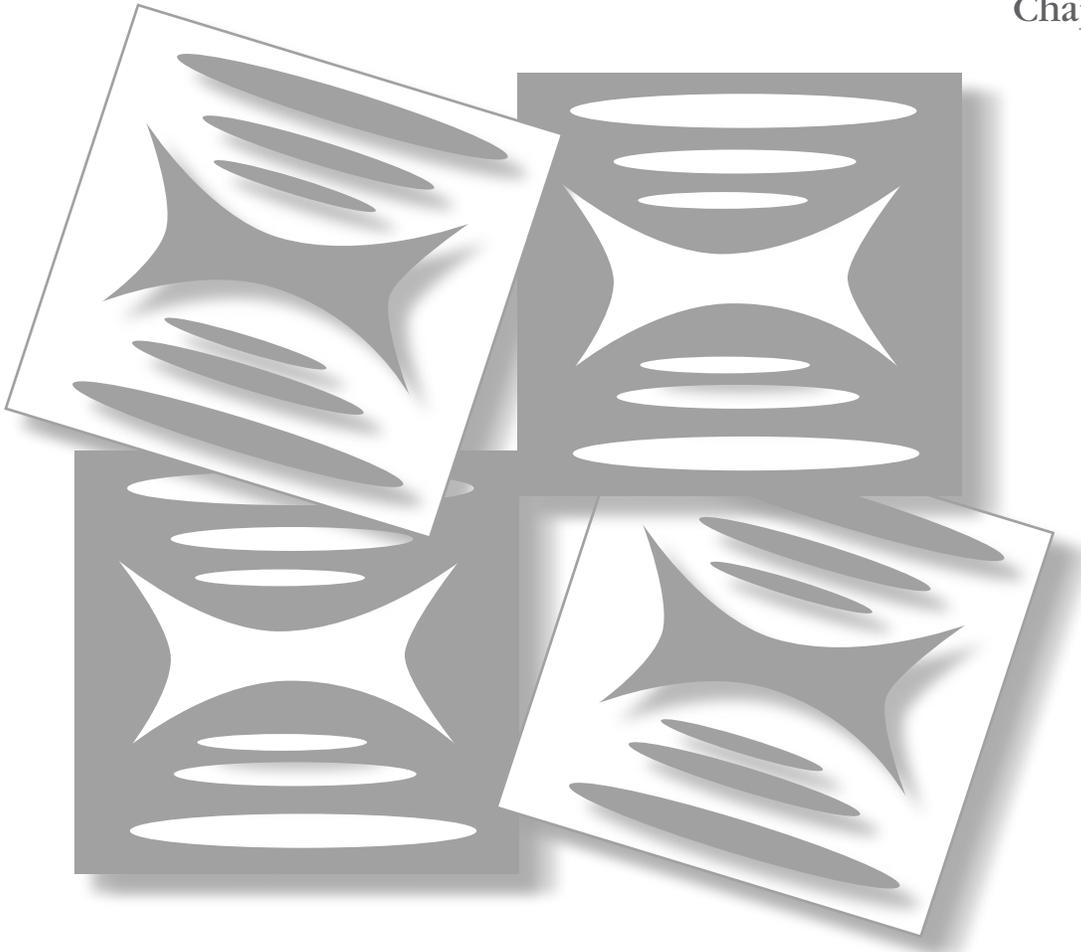
Further Reading:

ALNAP Training Modules for the Evaluation of Humanitarian Action *Module 3: Managing and Facilitating Evaluations of Humanitarian Action*, July 2003. <http://www.alnap.org/training.html>

The Evaluation Centre, Evaluation Management Checklists. <http://www.wmich.edu/evalctr/checklists/checklistmenu.htm#mgt>

Staff members who sit in on interviews or take part in the evaluator’s debriefs can hinder the openness of these discussions.

To avoid disappointment, organizations should be explicit about their expectations of the evaluation process.



**STAGE
THREE**

EVALUATION UTILIZATION

This chapter contains:

1. Three steps to using evaluation conclusions and recommendations
 - Reflect and Generalize
 - Apply
 - Share: Internally and Externally
2. Evaluation Utilization Checklist

INTRODUCTION

“Creative thinking may mean simply the realization that there is no particular virtue in doing things the way they always have been done.”

- RUDOLF FLESCH

In fact, the final report represents the completion of only the first two-thirds of a process whose main benefits are packed into the final third.

One of the most common misconceptions about evaluation is that the process is finished when the final report arrives on the project leader's desk. In fact, the final report represents the completion of only the first two-thirds of a process whose main benefits are packed into the final third. Imagine that the evaluators have brought a full bucket of water to the project leader on a scorching hot day. Someone has to reflect on how the water will be used and then take action to pour it into glasses, drink it, use it to water the plants, or bathe with it, otherwise it will just sit there and evaporate until it is no longer useful to anyone.

The latter third of the evaluation process focuses upon the use of the recommendations and conclusions. This stage is often referred to as the feedback loop, and it generally has two facets: internal learning and external sharing.

The benefits of utilizing evaluation results are multiple and they are at the heart of the effort to advance the peacebuilding field.

With the pressure to write proposals, implement projects and report to donors, internal learning may feel like an additional burden to the already overburdened practitioner. Yet the benefits of utilizing evaluation results are multiple and they are at the heart of the effort to advance the peacebuilding field. Within the project team, evaluation results enrich the discussion on whether the work is helping those it is supposed to help in the way it was planned. More pragmatically, it can indicate ways and means of improving the process. In addition, the advancement of models and designs indicates a level of expertise that can be compelling to donors.

As conflict transformation theory and practice continue to develop, a commitment by organizations to internal learning and external sharing has great potential to effect change within the practice of this field.

Sharing the evaluation results and the subsequent reflections and adaptations to the project externally increases transparency between the organization and its stakeholders and donors. Such sharing can also foster dialogue and reciprocity between organizations. It contributes to the growth and value of the field while reducing the amount of resources wasted through reinventing the wheel. As conflict transformation theory and practice continue to develop, a commitment by organizations to internal learning and external sharing has great potential to effect change within the practice of this field.

When do I start thinking about the “use” of an evaluation?

Begin thinking about how the evaluation will be used during the preparation stage. The evaluation objectives and how the project team hopes to use the information generated by the evaluation must be aligned. If, for instance, the evaluation objective is management and administration but the team hopes to use the evaluation information to understand more about the effects of the project, the team is unlikely to find the findings useful. External use of the evaluation should also be addressed briefly during the preparation stage when deciding on the evaluation’s audience and its potential readers.

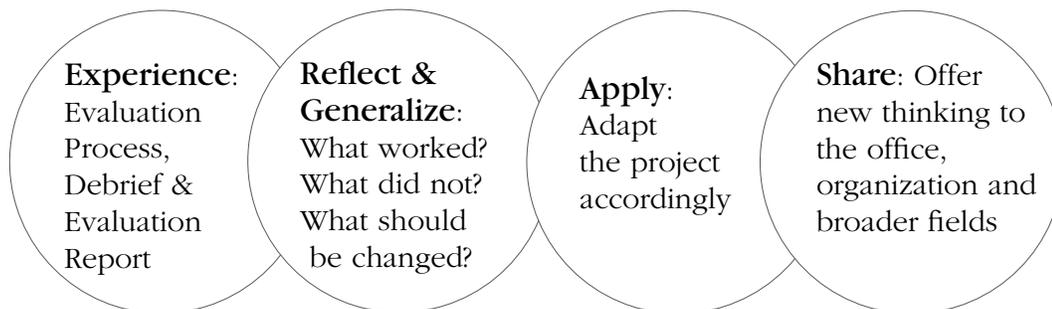
Of course, the evaluation implementation itself is a learning experience, but the real attention to internal learning starts when the evaluation team presents their draft conclusions while still in-country. A smart evaluation manager will remind the project team at this stage that there are more steps left to come.

The evaluation objectives and how the project team hopes to use the information generated by the evaluation must be aligned.

What are the steps needed to use the content of the evaluation report?

There are a series of overlapping steps, based on the Kolb adult learning stages outlined in the Learning Chapter page 6, that should be taken to maximize the internal and external benefits of an evaluation report. First, the project team goes through the experience of the evaluation – preparation, evaluation plan, process, debrief and review of the final report. The team then reflects on the conclusions and recommendations, and generalizes the information beyond the immediate project. From there the focus shifts to how this newly acquired knowledge should be applied to the project or beyond. Finally, the team considers what needs to be shared with other teams, with the broader organization, and with the field as a whole.

Applying Kolb’s Theory of Learning



As illustrated in the diagram, these steps overlap. What is important is that all the steps be taken and that they be completed in real-time. In other words, the process should not be delayed until the team thinks it will have more time to discuss the results.

Well-intentioned evaluation managers and project teams sometimes jump immediately to application. This is not altogether bad; however, these discussions often become stuck in the technicalities of the project and may not offer the opportunity to reflect on the macro-level issues or to address issues at a generic level separate from the immediate demands of the project.

EXPERIENCE: The evaluation has been planned, data collected, and tentative conclusions drawn. Both the in-country debriefing and the discussion with the project team have taken place. The remaining steps begin at this point.

REFLECT AND GENERALIZE: This step involves consideration of the findings from the perspective of the immediate project and long-term learning, review of the recommendations, and identification of the actions to be taken.

Once the evaluation is submitted, someone – often the evaluation manager – should be appointed as the learning facilitator. Remember, this role may not be necessary if it is already part of the evaluator’s responsibilities. The learning facilitator works with the project team to identify who should be included in these reflections. Together they should consider people located horizontally from the team as well as vertically above and below the team decision makers. Horizontally, staff members responsible for similar processes, partners, and stakeholders should be considered since they can benefit from the experience of others. Vertically, supervisor(s) – whether immediately above or even more senior – should also be included.

Next, develop a facilitated process to support the reflection and generalization. As with a conflict resolution process, simply gathering the right people is not enough. The process requires additional thought and creativity. For larger groups, this may involve a multi-stage process or different processes for different groupings. Sometimes it is helpful to focus solely on the evaluation conclusions in the first conversation and to leave the recommendations for the next discussion. This allows the project team to focus on the findings and to have a rare opportunity to discuss assumptions, principles and techniques outside the context of the day-to-day pressure of implementation.

In designing the process and selecting the participants, the conflict setting should not be ignored, particularly if staff members are from the area of conflict. Ensure that the process does not permit individual staff members, groups in the community or aspects of the governance structure to be blamed either for problems that arose or for a lack of results. Affirm the challenges of working on conflict issues and recognize the need to work around them.

As with a conflict resolution process, simply gathering the right people is not enough. The process requires additional thought and creativity.

In designing the process and selecting the participants, the conflict setting should not be ignored, particularly if staff members are from the area of conflict.

Whatever the process used, the discussion should facilitate reflection and generalization. Consider questions that range from the conceptual to the technical. The discussion need not be restricted to the evaluation results; however, be careful that the conversation does not avoid the thorny issues that an evaluation may raise. Depending on the evaluation objectives, the report conclusions may bring up some or all of the following questions for the project team to discuss:

- Did our theory(ies) of change work? Why or why not?
- Would a different theory of change have catalyzed a greater transformation?
- Was the context analysis accurate?
- Did our strategy link to the analysis in the best way?
- What happened as a result of these activities?
- What did not happen – that we expected to happen – as a result of these activities?
- Was the process we used within each activity the right one?
- How can we improve our techniques (e.g., training techniques)?
- What do the evaluation recommendations mean for the project and organization?
- What are the key learning points from this for the future?
- What were the critical factors for success?
- What are the pitfalls to look for next time?

To facilitate sharing and building institutional memory, the process and conclusions from these discussions should be documented.

Key Principles of Learning ²¹

- Help others actively interpret information so that they can construct new knowledge for themselves, rather than relying solely on paper formats.
- Situate abstract tasks in real contexts so that the relevance of the task is apparent and others can adopt the new knowledge.
- Separate knowledge from a single specific context in order to maximize knowledge sharing.
- Provide others with many examples of a new concept.
- Explain how essential features of a new concept are reflected in a variety of different settings.

²¹ Adapted from the *UNDP Handbook on Monitoring and Evaluation for Results*

For those recommendations not adopted, the reasons for not adopting them should be stated.

Identifying what did or did not work is an important first step. However, without an intentional discussion about what will be changed both in the immediate aftermath of the evaluation and in the longer term, it is unlikely that reflection alone will be enough to change the way projects are designed and implemented.

Begin by determining if everyone involved in the reflections conversation needs to be involved in setting action steps. This is facilitated by keeping the discussions on reflections and generalizations separate. Then review the team conclusions that resulted from the reflections and generalization discussion and determine the action steps to be taken to implement those conclusions. This discussion should also address which of the evaluation recommendations will be adopted. It is important to note that a project team is not required to act on every recommendation. The team should reflect, however, on all the recommendations and determine which will be accepted or rejected and why.

APPLY: This step moves the team back into action. To facilitate this, it may be helpful to develop a utilization plan. Action steps based on evaluator recommendations should be noted in conjunction with the recommendation from which they are taken. For those recommendations not adopted, the reasons for not adopting them should be stated. The utilization plan should also outline what will be done, by whom, and when.

This document can be a valuable addition to institutional memory since it captures decisions that may be useful for future evaluators, proposal writers, donors or new staff joining the team. Pending the content of the utilization plan, it can also be beneficial to attach it to the evaluation circulated to donors and key stakeholders. This shows that the organization is taking the evaluation seriously and has the capacity to learn and improve.

It may be helpful to extend the learning facilitator's role by a few months or even a year so that this person can check in with the team to ensure that the utilization plan has been implemented. As time passes, the learning facilitator can also initiate discussions on what difference those changes have made to continue the learning process.

SHARE: INTERNALLY Finally, what was learned needs to be shared with others both within the office and throughout the organization. At a minimum, circulate the evaluation and utilization plan to relevant parties such as department managers, senior personnel, field directors, etc. Other actions may include:

- Circulating the evaluation and utilization plan to wider range of staff

- Hosting a workshop on the results and lessons learned
- Adding panels to regional staff meetings on lessons from the evaluation
- Developing a short lessons learned summary
- Contributing a paragraph on the two critical factors for success in the internal newsletter
- Convening meetings to share the learning with regional or technical staff who support a wide range of programs
- Posting the evaluation on the organization intranet

Consideration should be given to the process used to share information if the goal is not only to assist people in knowing it but also applying the new information. Use the Key Principles of Learning depicted on page 182 to inform design of these mechanisms.



“I can honestly say that not a day goes by when we don’t use those evaluations in one way or another.”

Written by *M. M. Rogers* and illustrated by *Lawson Sworb*

Within the fast-paced context of a conflict situation, engaging in systematic reflection can prove to be extremely challenging as the needs or opportunities that prompted the project in the first place may no longer be relevant.

SHARE: EXTERNALLY In addition to the learning benefits from the evaluation results, a wise organization reaps an array of valuable benefits from distributing the evaluation externally. Different formats, such as key results document, may need to be developed for different external audiences in order to be effective. Possible external uses include:

Donor Relations

- Incorporate evidence-based results to strengthen proposals
- Use the document as a means of outreach to new donors
- Strengthen established relations through discussions about the results and the changes the organization is making as a consequence of the evaluation

Public Relations

- Host a meeting with interested stakeholders to discuss the results
- Include key quotes on websites and brochures
- Add a “Results” or “Accomplishments” section to the annual report
- Include the results in a key talking points packet for the press
- Produce an “Accomplishments” brochure that focuses on the difference the work has made in the world
- Write a concise summary and circulate to electronically to relevant peacebuilding listservs

Academia

- Write journal articles that include the evaluation results or that are based on the evaluation experience
- Invite academics and researchers to use the data collected for the evaluation in their studies
- Present papers on the evaluation at conferences

I’m really busy. Can all of this be put together in a checklist?

Within the fast-paced context of a conflict situation, engaging in systematic reflection can prove to be extremely challenging as the needs or opportunities that prompted the project in the first place may no longer be relevant. In addition, a sense of urgency can develop that one needs to learn new things to address up-and-coming problems. If this sense dominates, the ability of the peacebuilding field to improve

and to increase its effectiveness in achieving its goal of building positive peace around the world will remain limited. Moreover, project teams and organizations will not learn to do their work better, which will also limit the sharing of knowledge within the field.

Use the following Evaluation Utilization Checklist as a tool to ensure that none of the steps are missed. The “Lead Actor” column indicates who is responsible for the task, while the “Who is involved?” column covers key participants within the organization. The “When” column requires a sense of timing for the task.

Evaluation Utilization Checklist

| Tasks | Lead Actor | Who is involved? | When |
|--|-------------------|-------------------------|-------------|
| Determine evaluation objective(s) | | | |
| Conduct draft conclusions and debrief with project team | | | |
| Determine who will be involved in reflections conversation | | | |
| Develop process for reflections conversation | | | |
| Circulate final evaluation report to all involved in conversations | | | |
| Conduct first reflections conversation | | | |
| Document thoughts and ideas from conversation | | | |
| Determine who should be involved in making decisions about changes at the project, program and/or organizational level | | | |
| Identify adaptations to be made including responses to the evaluation recommendations | | | |
| Development of utilization plan | | | |
| Evaluation and utilization plan (including new knowledge) circulated to relevant staff | | | |
| Monitor how the new learning and utilization plan has been applied | | | |
| Incorporate results into organizational working knowledge through a variety of forums such as <ul style="list-style-type: none"> • Workshops • Internal Newsletter • Panels • Lessons Learned Briefing | | | |
| Consider use of results in donor relations, public relations and academia | | | |
| Circulate to other organizations in the field | | | |

Utilization from a long-term perspective

Over time, an organization will accumulate more and more evaluations, which will create opportunities for greater utilization, improvement in evaluation quality, and the maximization of learning. Some options to consider include:

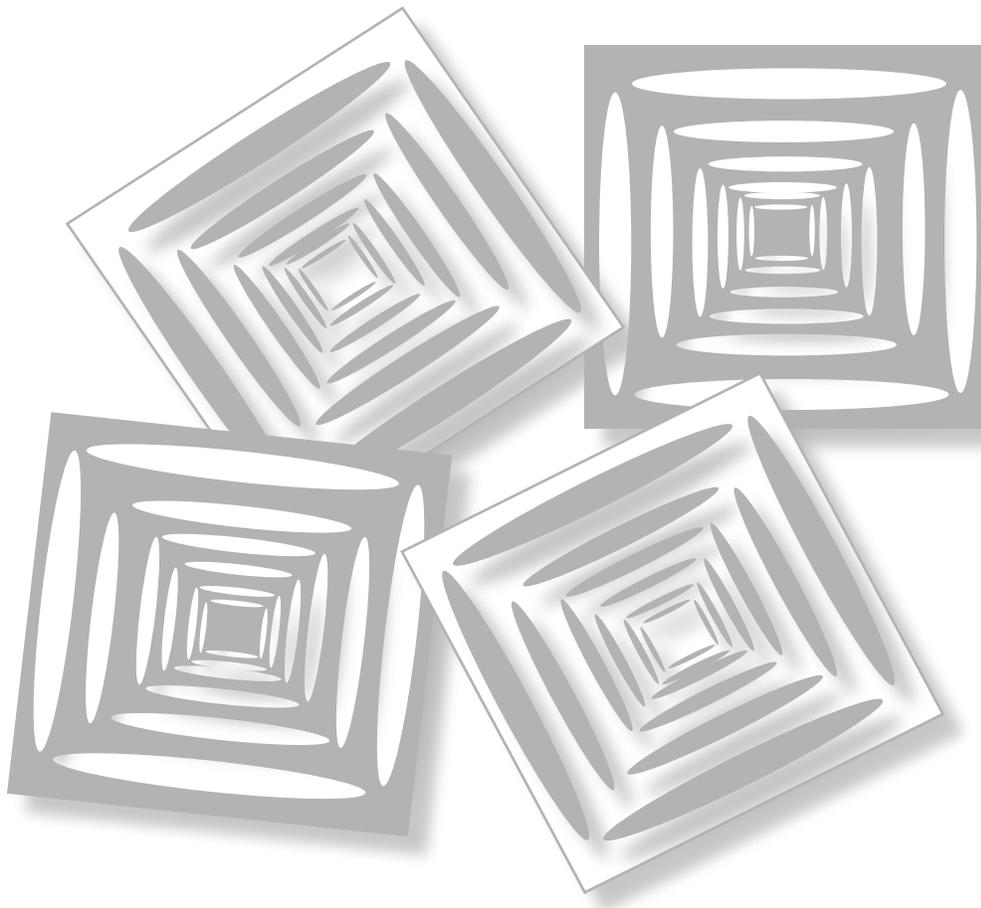
- **Evaluation Synthesis:** *The gathering and summarizing of the results of evaluation studies of similar programs (by theme or tool).*
- **Standards of Practice Guidelines:** *Development of standards of practice or guidelines that evaluations can use to assess programs against.*
- **Meta-evaluation:** *Assessing existing evaluations against pertinent standards to help improve evaluation implementation in the future. This is a form of evaluation quality control.*

Further Reading:

Michael Quinn Patton, *Utilization-Focused Evaluation*, Third Edition, SAGE Publications, 1996.

David A Kolb, *Kolb's Theory of Adult Learning*, 1984.

<http://www.learningandteaching.info/learning/experience.htm>



In Design, Monitoring And Evaluation For Conflict Transformation **ETHICS**

This section includes:

1. Discussion of unethical practices
2. Categories and strategies for dealing with common ethical issues
 - *Protection of people*
 - *Freedom from political interference*
 - *Quality data collection techniques*
3. What is informed consent?
4. Internal versus external evaluators and ethics

INTRODUCTION

“To enjoy the things we ought and to hate the things we ought has the greatest bearing on excellence of character.”

- ARISTOTLE

This chapter discusses ethical issues in design, monitoring, and evaluation of peacebuilding projects. It begins by offering guidance on what constitutes an unethical practice and how such practices might arise at each stage of the project cycle. Not surprisingly, the ethical issues related to baselines and evaluations overlap in a number of areas, whereas the design stage has several unique ethical challenges of its own. Within each section, practical strategies for preventing and avoiding unethical choices are offered. The practice of informed consent is considered in relation to the realities of evaluation of peacebuilding. Finally, the chapter explores the different ethical issues faced by internal and external evaluators.

What is an unethical practice?

Not knowing what constitutes best practice is incompetence. Knowing what best practice is, but not knowing how to achieve it, may be inexperience. Knowingly not following best practices, when one knows how to achieve it, is unethical.²²

There are many types of unethical practice in design, monitoring, and evaluation of conflict transformation programs. On one hand, there are ethical issues that commonly occur in DM&E of peacebuilding that generally have a “right” and a “wrong” answer. Changing data to represent a project in a more positive or negative light, for instance, is clearly wrong.

Conversely, an ethical dilemma often does not have a clear right or wrong answer, and because of this ambiguity, decisions should be considered on a case-by-case basis. The cultural norms, values, and experiences of those involved often play a significant role in the decision. An example of such an ethical dilemma can be seen in the decision about whether or not to insist on the equality of youth voices in decisionmaking as part of a participatory evaluation in a society that honors elders and the roles they play as key decisionmakers.

For some, what is discussed in this chapter may be viewed either as standard political actions to benefit one side or another or simply the lack

²² Nick Smith, *An Analysis of Ethical Challenges in Evaluation*, American Journal of Evaluation, Volume 23, No. 2(2002)

of application of best practices. Regardless of the label applied, if the action does not fall within appropriate principles of conduct for design, monitoring, and evaluation, it fits within this discussion on ethics.

What are the ethical challenges common in the design stage?

When discussing the development of a peacebuilding project, innumerable ethical challenges may arise. The four challenges discussed below, however, are those that are likely to apply in most situations.

First, practitioners have an ethical obligation to involve the parties to a conflict in determining the changes that will satisfy their respective interests. Although people in conflict may not always be able see how to transform their disputes at the outset of discussions, they ultimately make the final choices once additional options have been explored.

Second, practitioners have an ethical responsibility to consider any possible negative ramifications that may occur as a result of a project and to do everything feasible to eliminate them. Consider, for instance, a youth project in Israel that targeted teenage boys, ages 13-15, from politically hard-line families. The goal of the project was to change attitudes from supporting violence to a recognition that there are multiple ways beyond violence to resolve the political situation. As the project progressed, some of the participants started to challenge the adult members of their families in political discussions. In one case, this led a father to physically assault his son as punishment for what the father saw as the son's disloyal and disrespectful opinions. Such an unintended negative effect might have been prevented if, in the design stage of the project, this scenario had been identified and preventive measures adopted. These measures could have included engagement with parents or the incorporation of techniques for dealing constructively with families about sensitive issues so that the participants would be prepared for such a situation.

Third, practitioners have an ethical obligation to develop projects that maximize the opportunities for change. This maximization is determined on a situation-by-situation basis, but it broadly encompasses creating change among the most people, in the fastest way possible, for the greatest possible positive change, and with the least possible negative consequences. The ethical challenge arises when project designs that do not maximize the opportunities for change are seen as easier to implement or have more readily available sources of funding.

Finally, the development of indicators can be an ethical issue. For example, there may be indicators that reflect changes of less importance or that signal changes on issues that are not directly affected by the project but which present the work in a more positive light than would an accurate indicator.

In an attempt to address these ethical challenges, organizations can add a set of questions to their project design processes that they routinely review. These questions might include some of the following:

- If there were absolutely no restrictions in terms of capacity, time, or funds, how would we modify this project?
- Have discussions with the prospective donor taken place to explore options that may be more productive or beneficial to the stakeholders?
- Have the connections between the analysis and the proposed project been explicitly outlined?
- Can the stakeholder's perspective be seen in the final design?
- What are the potential negative results that could occur as a result of this project? What steps need to be taken to minimize the potential negative consequences to participants, staff, or the community?
- Are there other programs currently operating to which this project should be connected in order to maximize results?
- If the team members could only implement one project, which would they select as the most important, and would it make a difference?
- Were other options fully discussed based on the conflict assessment, particularly those not part of our regular activities?

What are some of the common ethical challenges for baselines and evaluations?

There are a number of ethical issues and dilemmas to consider when implementing baselines and evaluations. In some cases, the same issues apply to monitoring as well. The ethical challenges can be grouped into three broad categories: protection of people, freedom from political interference, and quality data collection techniques.

1. Protection of People

The ethical challenges related to the protection of people can be subdivided into six major themes: avoiding personal duress, guaranteeing confidentiality, considering safety, setting realistic expectations, protecting the organization's credibility, and avoiding research subject fatigue.

This category (protection of people) spans all units of analysis, from the individual, to the implementing organization, to the target group as a whole (e.g., all Hutus or all ex-combatants).

AVOIDING PERSONAL DURESS: Data collectors should consider the potential negative consequences that could arise from delving into an individual's personal experience. For instance, silence can be a coping strategy for some victims of violent conflict; however, they may be asked to talk about their experiences as part of the evaluation process. This process therefore risks undermining the participants' coping strategy without offering the necessary support structure to provide assistance if it is needed.²³

Evaluators should approach some groups in places of war, such as victims of rape or torture, with caution, and ideally consult with experts on the appropriate ways to engage with these groups, if at all. However, conflict zones are rife with individuals who have unhealed psychological wounds and trauma that are not apparent. Evaluators should therefore look for signals of duress, such as agitation or tears, in their subjects and be prepared to handle the situation appropriately. Seeking advice from experts on this issue, prior to data collection, is a prudent step for the professional evaluator.

GUARANTEEING CONFIDENTIALITY: It is important for individuals providing data for a baseline or evaluation, whether through surveys or in one-on-one interviews, to understand how their names will be used in connection with the information they provide. The evaluator must explain clearly how the information will be attributed in the final deliverable. In other words, will the person's name be used, along with her/his ideas, in a quote format or will attributes be used to provide a context for the comments (e.g., women in the village), or will the information simply stand alone?

In conflict settings, where speaking out against one's group or the government, for instance, may prove deadly, the norm is to guarantee confidentiality to all individuals who participate. In this case, not only does the evaluator need to explain that the data is confidential to each individual, she/he must do preparatory work to ensure that confidentiality can be guaranteed. More on this issue can be found later in the chapter under Informed Consent, page 198.

Special care is due when writing the baseline or evaluation report once confidentiality has been promised. In local settings or where people are assumed to hold particular views, even general attributes in connection with specific statements may be identifiable by the community. For instance, if there are only ten positions on the district council and six individuals have held their positions for years, attributing a statement to a new member to the district council is almost the same as using the

²³ Cheyanne Church and Julie Shouldice, *Part II*.

person's name. This can become a difficult dilemma when the identity, position, or standing of a person inherently provides insights on her/his statements or opinions.

CONSIDERING SAFETY: In conflicts where communities are segregated, like Kosovo or Northern Ireland, being seen speaking to an outsider can be enough to cause suspicion within an individual's community. A number of questions might result: Who was the stranger? Why were they here? What did they want to know? What did you tell them? In a calm situation, suspicion may end with harmless gossip; however, if tensions rise or are already high, the suspicions could grow into more serious outcomes for the individual, such as expulsion or physical harm. Evaluators have an ethical responsibility both to consider the safety of the individuals who provide them with information and to plan their data collection efforts to minimize any possible risk. Where meetings take place, who introduces the evaluator to the individual, and who should be told about the evaluation and the purpose of the visits are all important considerations for an ethical evaluator.

Another safety dilemma can arise when an evaluation team hires members from the conflict setting. The members from the community may have greater freedom of movement in areas experiencing active conflict and often travel alone to these locations to collect data. What is the team's responsibility for the personal safety of these local team members when they enter high-risk areas? The dilemma lies in what constitutes too much risk. It may be useful to consider the following rule of thumb: If the team member from the area would not travel to the conflict area independently, regardless of the foreign vehicle or official trappings, other options should be considered for accessing the data being sought. (See the Evaluation Management chapter, page 168, for more information on accessing data in situations that are too dangerous to enter.) In addition, do not assume that a team member from the area is aware of the security concerns at the time the work is to be done.

SETTING REALISTIC EXPECTATIONS: Anyone collecting data for a baseline, monitoring, or evaluation should be careful not to set undue expectations during the data collection process. It is often the case that, in an effort to express thanks, generate excitement, or convince people to answer questions, the data collector inadvertently raises expectations unrealistically. Consider the following example.

A practitioner was monitoring the progress toward results of a program seeking to increase a community's knowledge of the city's grievance procedures. He wanted to speak to community members who were residents of government-subsidized housing in an area of violent crime because the project team felt that the data should be disaggregated for socio-economic standing. (See the Methods chapter, page 216, for more information on disaggregated data.)

Convincing residents of the housing project to speak with him was difficult, however. In an effort to entice more people to cooperate, he opened the conversation with the following statement: “If our group is going to be able to help people deal with crime in this area, we need people to answer a few questions.” This led many of the residents to conclude that the NGO was going to work actively in their community to decrease crime. Arguably, knowing more about grievance procedures might help in this regard; however, the NGO’s intention was never to directly address crime and violence. As a result, it set unrealistic expectations.

PROTECTING THE ORGANIZATION’S CREDIBILITY: Those collecting data should also be aware that their actions are often deemed to be an extension of the organization being evaluated. Consequently, if a team member behaves inappropriately it can harm the organization’s reputation, and – in more serious cases – the inappropriate behavior may derail any progress achieved to date from the work.

Consider the following example from Northern Ireland. “[A]n evaluator... enters a tense conflict situation to evaluate a cross-community dialogue project with leaders of opposing communities. The evaluator is permitted to meet with the participants in the program because of the goodwill and trust established between the conflict parties and the implementing agency. However, if the evaluator does not operate within the norms of the communication established by the agency such as meeting with an equal number of representatives from each side, or is interpreted as being biased by one of the parties, this can severely damage the agency’s credibility with the parties and constrain the dialogue process.”²⁴

AVOIDING RESEARCH SUBJECT FATIGUE: In areas where a great deal of research is done, there can be problems with beleaguered research subjects being asked repeatedly to offer information on similar themes in relatively short periods of time. Not only does such repetition steal valuable time from the individual, it also dilutes the authenticity of the answer since the person has been asked about the same subject so many times that her/his response becomes almost “pre-recorded.”

Organizations considering an evaluation would therefore be prudent to ask other agencies working in the area or their donor if other evaluations are pending. Sometimes it is possible to combine evaluation or baseline projects. This not only shows respect for the research subjects but it can also decrease the cost of the research to the organization. Furthermore, professional evaluators should inquire about the possibility of cooperation with other organizations at the earliest stage possible within the evaluation since there may still be time to combine research with other efforts.

²⁴ Church and Shouldice, *Part II*.

2. Freedom from Political Interference

Evaluations should be conducted free from political interference from the implementing organization, the donor, the evaluation team, and the stakeholders. There are many ways, and many different reasons why, political pressure might be applied to an evaluation process. Implementing organizations may see the evaluation as a way to promote themselves to donors and, as a result, will steer the evaluator toward only those people who will speak positively about the organization. Donors may view an evaluation as a way to justify a decision to end funding to a sector or organization by requiring a methodology that misses many of the positive results. The evaluation team may wish to secure ongoing contracts with the implementing agency by presenting the agency in an unearned positive light. Stakeholders may see the evaluation as the only way to access additional resources for their community and, therefore, they may lobby the evaluator to make specific recommendations.

Evaluations provide far more opportunity for political interference than do baselines, though baselines are not exempt from meddling. Some of the more common political interferences can be found in the table below. An “X” indicates whether the political interference applies most commonly to baselines, evaluations, or both.

Common Political Interference in Baseines and Evaluations

| Common Political Interference | Baseline | Evaluation |
|--|-----------------|-------------------|
| Implementing agency pressures the evaluation team to omit weaknesses from the final evaluation report. | | X |
| Staff members ask the evaluation team to show them in a positive light. | | X |
| Donors require the evaluator to use a methodology that is not optimal for the information being sought. | X | X |
| Staff pressure participants into being part of the study. | X | X |
| The implementing agency or donor already has an answer and writes the terms of reference in a way that sets up the evaluator to justify that answer. | | X |
| Staff members coach project participants on the kinds of responses they want given to the evaluator. | | X |
| Evaluators are pushed toward specific sets of people who are unusually positive or negative. | X | X |
| Not providing the evaluation team with reports that capture concerns or negative effects of the project. | | X |
| Not including on participant lists those who have dropped out of the project. | | X |
| Creating documents such as reports or project logs to meet the evaluation team request during the evaluation. | | X |

There are a number of strategies that can be adopted when there is political interference. Some of the most common strategies recommended by professional evaluators include:

FRAMING THE ISSUE OF CONCERN: Consider the attempted interference as a regular part of negotiations rather than as an unmanageable impediment to a quality evaluation process.²⁵ In this case, the evaluator should reframe the concern as an issue that requires additional negotiation with the party exerting the political interference. This technique is often combined with Communication & Education, which is described next.

COMMUNICATION & EDUCATION: Ensuring that all stakeholders understand the steps in the process and the rationale behind them will decrease opportunities for misunderstanding and potential malpractice. In the first stage of the evaluation, explicitly develop the principle of transparent communication, whereby all parties provide explanations of their requests and choices, and provide opportunities for discussion.

DETAILED & DOCUMENTED PLANNING: An evaluation plan and terms of reference that are well-thought through and documented can be very helpful for at least two reasons. First, they offer clear boundaries and decisions that are less open to interpretation. Second, they provide an historical reference that the evaluator can refer to later if inappropriate pressure arises during the evaluation process. A detailed contract with clearly defined grievance procedures can also be helpful if disagreements arise.

TIMELINESS: When actions occur or statements are made that seem intended to exert undue political pressure, they need to be faced immediately and directly.

INCREASE THE SEATS AT THE TABLE: The more stakeholders represented at the table, the more likely political interference will either not arise or will be handled in an appropriate manner.

CHECKS & BALANCES: Having an evaluation team rather than an individual evaluator can provide checks and balances when ethical challenges arise. In addition, if political pressure is at the heart of an issue, there is always more strength in numbers (i.e., as opposed to an individual evaluator on her/his own). Another good check and balance to put in place is an independent evaluation manager. See the Evaluation Management chapter for more information on the role of the evaluation manager, page 137.

CONSULT EXPERTS: If there is any uncertainty as to what is acceptable, consult an expert. If the organization has internal DM&E expertise, check to see if there are any norms that the organization has chosen

²⁵ Nick Smith, *An Analysis of Ethical Challenges in Evaluation*, American Journal of Evaluation, Volume 23, No. 2 (2002).

to utilize or request guidance. If there are no internal resources, use evaluation network listservs to request input or contact academics in the evaluation field. If the question is brief, people are generally happy to provide their input.

ADHERENCE TO AND DISCUSSION ABOUT PROFESSIONAL PRINCIPLES: Professional associations increasingly issue principles or norms of ethical practice. National evaluation associations are no different; therefore, be sure to check if professional principles have been issued for the country in which the evaluation will be conducted. If none exist and/or there is no national evaluation association, a good alternative is the American Evaluation Association Guiding Principles below. These principles may need to be adapted somewhat to other settings, but they offer a useful point of reference for a discussion with stakeholders in an evaluation.

American Evaluation Association Guiding Principles

- A. Systematic Inquiry: Evaluators conduct systematic, data-based inquiries about whatever is being evaluated.
- B. Competence: Evaluators provide competent performance to stakeholders.
- C. Integrity/Honesty: Evaluators ensure the honesty and integrity of the entire evaluation process.
- D. Respect for People: Evaluators respect the security, dignity and self-worth of the respondents, program participants, clients, and other stakeholders with whom they interact.
- E. Responsibilities for General and Public Welfare: Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare.

HONOR YOUR OWN INTEGRITY: If you feel you are being asked to do something that does not “intuitively” feel right, raise the issue with the organization before you proceed. Some may call this the “Can I sleep at night?” measure.

3. Quality Data Collection Techniques

In addition to using best practice in data collection to ensure sound and credible inputs for analysis, evaluators also need to consider some ethical challenges that can affect data quality.

One such challenge is whether or not to reimburse people for the time they have given to provide information. In many peacebuilding programs, the average participant would qualify as being a member of the

world's poor. Taking two hours of time to participate in a focus group rather than earn money or gather food may have a substantial effect on the person's livelihood. Should they be reimbursed for their time, with money, transport, or food? Classic social science research would state that they should not be reimbursed since remunerating people for their opinions may cause them to alter their statements or responses to be unduly positive or supportive of the topic.

However, the rules of social science methodology were created in the "developed West" and often need to be adapted to non-Western contexts. If possible, it is always better to avoid providing material incentives since they can potentially skew the results. Nonetheless, this is not always the most ethical stance and, in certain cases, creatively identifying ways to reimburse people for their time is appropriate. Offering lunch, funds for transport, or a small item such as a bucket are potential examples. If such items are offered, assure each person that the reimbursement is guaranteed regardless of the information offered.

The second dilemma to consider is the balance between respecting local customs and advancing an agenda the organization or evaluation team deems important to the project. One of the most widely known illustrations of this is gender inclusion. It is commonplace for evaluators to want to engage men and women in their data collection, yet in some situations, accessing women's opinions may be counter to local customs. This can be particularly true if the evaluators are solely male and wish to speak to women without the presence of local men. Should the beliefs of the evaluation team override local customs? In these situations, it may be best to turn to the implementing organization for guidance.

What is Informed Consent?

Informed consent is the process of educating participants in the research about the purpose, procedures, risks, benefits, and alternatives to participation. In social science research at the academic level, informed consent is a standard and required part of any research project. It is seen as an ethical obligation of the researcher and as a key part of the protection of the people involved in the study. In these cases, consent needs to be obtained in written form from participants before they become involved in the research. It is far more than simply obtaining a signature on the consent document. It is about the individual's understanding and willingness to participate in the study.²⁶

This standard of written consent is not yet the norm in international conflict transformation evaluation. Complying with the written documentation requirement may never be feasible because of illiteracy as well as confidentiality and security concerns in conflict settings. The essence of informed consent holds true, however, regardless of

²⁶ Informed Consent Overview, Institutional Review Board, University of Minnesota, 1998 <http://www.research.umn.edu/consent/modlsoc/modlsec1.html>

the setting. Participants in an evaluation should be informed of the purpose, process, risks, and benefits of participation and be given the opportunity to decline to participate.

Are there different ethical dilemmas and issues for internal versus external evaluators?

Ethical challenges do not generally differ between internal and external evaluators. What does seem to differ is the way in which issues are resolved. One of the primary factors behind this difference is the different relationship to the organizational structure that each holds.

It is the nature of those relationships that is critical for considering ethical dilemmas. Internal evaluators are situated directly within the organization whereas externals are outside the organization and are related to many different entities at the same time. The organization that constitutes the most important relationship to an external evaluator is rarely the one she/he is evaluating. Yet for the internal evaluator, the focus of the evaluation – a project within her/his organization – is generally the most important relationship to the evaluator's professional position.

Ethical dilemmas therefore arise with people who the internal evaluator knows well and works for routinely. The internal evaluator often feels that she/he has fewer options in challenging situations. To foster a sense of belonging and long-term community, an internal evaluator may feel the need to be more conciliatory about challenging issues. External evaluators generally have more latitude because their connections to the group involve a particular project and they have been brought in for their expertise on that project.

This situation may mean internal evaluators are more vulnerable to poor practices exerted by the organization or donor which result from conflicting roles associated with being both a professional evaluator and a member of an organization. However, the personal relationships that consulting professionals develop with their clients, and the expectations engendered by clients' direct hiring and reimbursement of the professional, may also exacerbate ethical dilemmas. Due to the inherent power dynamic, it can appear against the consulting professional's best interest to pursue ethical norms that seem to conflict with a client's self-interest.

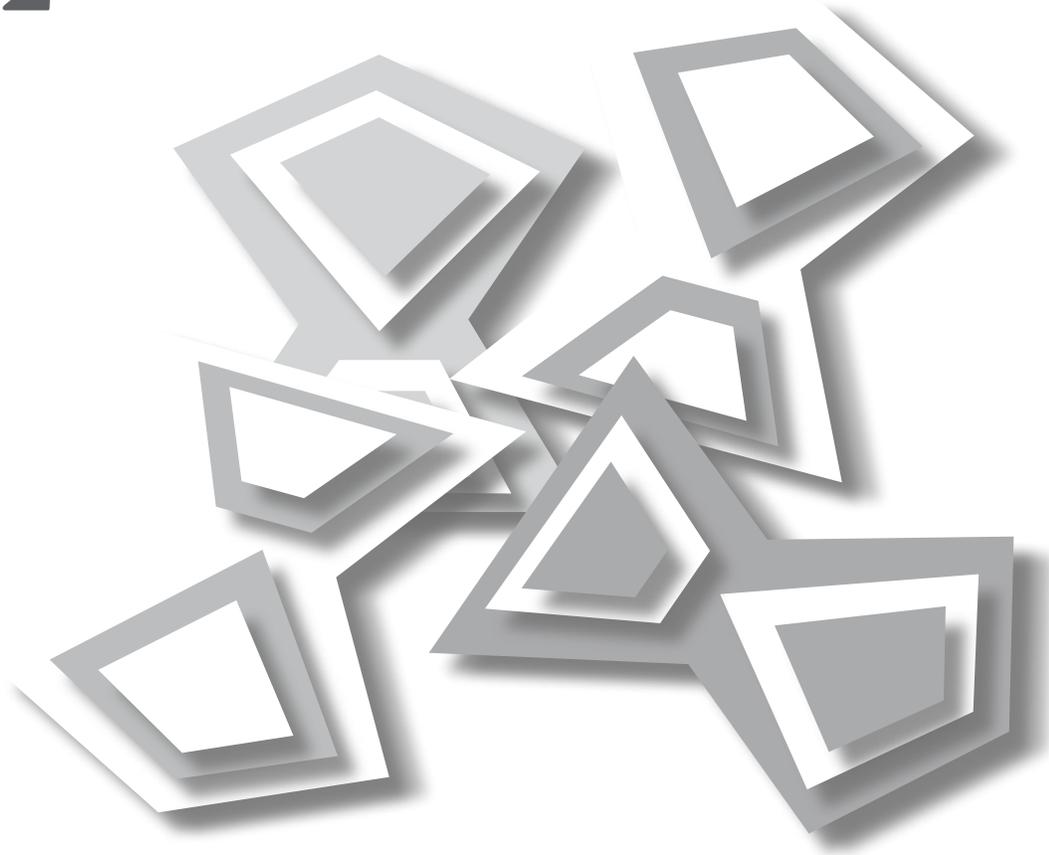
Further Reading

Brian English, "Conducting Ethical Evaluations with Disadvantaged and Minority Target Groups," *Evaluation Practice*, Volume 18, Number 1, 1997

Jody L. Fitzpatrick and Michael Morris (eds.), "Current and Emerging Ethical Issues in Evaluation," *New Directions for Evaluation*, Number 82, Summer 1999

Ethics, Canadian Evaluation Society <http://evaluationcanada.ca>

Guiding Principles, American Evaluation Association <http://www.eval.org/Guiding%20Principles.htm>



METHODS

This chapter contains:

1. Overview of basic concepts
2. Standard data collection methods
3. Cautionary note about data collection
4. Guiding questions to aid in selecting methods
5. Instrument development and testing
6. Disaggregated data
7. Data analysis
8. Unique peacebuilding tools
9. Sound basis for generalization
10. Record maintenance systems

INTRODUCTION

“All men by nature desire knowledge.”

-ARISTOTLE

This chapter introduces peacebuilding practitioners to the options and considerations for selecting the means of data collection for baselines, monitoring, and evaluation. In logical frameworks, methods are referred to as means of verification (MOV), while in other circles they are called research instruments. In this chapter, the term “research” will be used to mean any data collection done for baselines, monitoring or evaluation.

Data collection methods have been developed over a period of many decades and now have well-established standards and techniques. A comprehensive introduction to methods requires a manual or two of its own. For the purposes of the practitioner, one needs to understand the core concepts and terminology as well as when to use which method for the best results. The techniques of designing and implementing those methods are beyond the immediate needs of the average practitioner, hence the scope of this manual.

It should be made clear what this chapter is not. This chapter is not trying to make researchers out of practitioners. It does not have enough information on how to select, design, and implement methods nor does it cover how to analyze the resulting data such that a beginner could do so effectively. Instead, this chapter is intended to prepare a practitioner to have knowledgeable conversations with professional evaluators in order to make informed choices.

What are the basic concepts I need to know about data collection?

Mastering a few core concepts and their associated terminology is the first step in understanding data collection.

Key Terms

Fundamental to research is the notion that data and conclusions are two different things. Data is the building block of information and is often thought of as statistical or quantitative, although it may take many other forms, such as transcripts of interviews, maps, photographs or videotapes of social interactions. Conclusions are drawn from data through analysis.

Data collection methods have been developed over a period of many decades and now have well-established standards and techniques.

Fundamental to research is the notion that data and conclusions are two different things.

Methods are the means of acquiring the necessary data.

Methods are the means of acquiring the necessary data. Once gathered, the data is analyzed in order to generate conclusions.

There are a number of other terms used in research that are important to understand. These terms are listed below in the Methods Key Terminology table, and their accompanying definitions have been made informal and simple.

Methods Key Terminology

| | |
|---------------------------|---|
| Bias | To be inclined toward a particular way of looking at or understanding something |
| Causality | The direct effect of one event on a future event |
| Correlation | The extent to which two or more things are related to one another |
| Disaggregate | To separate into component parts |
| Generalization | The extent to which one can come to broad conclusions about a group or phenomenon based on information gathered from a set of representatives of that group or phenomenon |
| Reliability | Do repeated applications of the method, even when different people apply it, result in the same outcome? |
| Sample | Representative members of the entire client base for the activity |
| Statistically Significant | Meaningful, measurable relationship or level of change |
| Unit of Analysis | The primary entity under evaluation, such as individuals, groups, artifacts, geographic units or social interactions |
| Validity | Does the method measure what it is supposed to? |

Quantitative & Qualitative: Methods and Data

Quantitative and qualitative are important concepts for practitioners to understand. Quantitative methods are used to gather data to be analysed in numerical form. They pose questions of who, what, when, where, how much, how many, and how, generally in the form of surveys and questionnaires. These methods are designed to produce data that tells us how many people do or think something, and which is statistically reliable. Quantitative data typically is in numerical form such as averages, ratios or ranges.

Qualitative methods have greater flexibility and pose questions in a more open-ended manner. They give an in-depth understanding of why people hold particular views. They also explore how people make judgments, in a way that structured quantitative research cannot. Qualitative

Quantitative methods are used to gather data to be analysed in numerical form.

methods are not intended to be statistically reliable, but findings can — if participants (those who provide data to the study) are broadly representative — be strongly indicative of the population as a whole. Standard qualitative methods include interviews and focus groups. Qualitative data is typically words or text, though it can include photographs, video, or sound recordings.

The descriptions provided thus far cover quantitative and qualitative data and data collection methods. It is important to understand, however, that there are also quantitative and qualitative analysis approaches. See the question about analysis on page 50 for more information.

There has been an ongoing debate about which method — quantitative or qualitative — is better suited for baseline, monitoring, and evaluation purposes. Advocates for a quantitative approach argue that their data is hard, rigorous, credible, and scientific. On the other hand, proponents of the qualitative method contend that their data is sensitive, nuanced, detailed, and contextual. For peacebuilding baseline, monitoring, and evaluation purposes, this debate is needless because both approaches are necessary.

Mixed methods — the use of both qualitative and quantitative approaches — are becoming the new norm because they produce a richer set of information to meet the needs of conflict transformation projects. Numbers alone rarely answer the questions that peacebuilders have regarding why and how social change occurred. Perceptions and feelings that cannot be generalized to a greater population do not provide the complete picture either, necessitating the use of quantitative methods.

What are the standard data collection methods?

All of the standard social science methods, such as surveys or interviews, as well as participatory techniques like mapping, can be utilized in baseline, monitoring, and evaluation of conflict transformation. Each of these methods can be implemented individually or in combination with each other depending on the research needs. This chapter provides short overviews of some of the more commonly utilized methods and their strengths and weaknesses. How to choose the right method for research follows.

- **DIRECT OBSERVATION:** Watching, taking notes, and recording specific actions within a target community, such as communications, spatial interaction, or exclusion. The observation can be focused on a project process in which the people participating in the intervention are observed. It can also be focused on changes, such as in people's behaviors and attitudes, which involves watching people go about their daily business at home, in the community, or in the fields.

Quantitative methods are used to gather data to be analysed in numerical form.

Qualitative methods have greater flexibility and pose questions in a more open-ended manner. They give an in-depth understanding of why people hold particular views.

All of the standard social science methods can be utilized in baseline, monitoring, and evaluation of conflict transformation.

- **INTERVIEWS:** One-on-one contact with stakeholders, either in person or by telephone. These can be formal structured exercises, where a strict interview protocol is followed, or semi-structured meetings that are partially structured by a flexible interview guide. For comparability purposes, a minimum degree of commonality must exist in unstructured interviews.
- **FOCUS GROUPS:** Small-group conversations that seek to understand how people feel or think about an issue, product, service or idea. Focus groups have a specific purpose, size, composition, and process. They are best conducted with 6-8 people who are selected because they have something in common. Leading focus groups requires a skilled moderator and is best done in a comfortable, permissive environment. Such groups are a compromise between participant observation and more in-depth interviews.
- **PARTICIPANT DIARIES:** These are narrative descriptions of a personal experience. They may be open-ended to allow individuals to capture what was of importance to them each day or week, for example. Participant diaries can also be structured so that individuals take note of specific attitudes, events, behaviors that they have experienced in the allocated timeframe (e.g., daily, weekly).
- **PHOTOGRAPHY/VIDEO:** Utilizing photographs or video to collect visually represented information
- **PROJECT DOCUMENT REVIEW:** Collecting, reading, collating, and analyzing key documents such as proposals, donor reports, annual reports, case studies, etc.
- **QUESTIONNAIRE:** A set of specific, targeted questions to which stakeholders respond in writing. The questions must reflect cultural awareness and be language sensitive in addition to fitting within a set of formal methodological standards. Questionnaires can be distributed electronically, by post or by hand. For a statistically significant conclusion, the number of responses needed depends on the total population size. However, to do statistical applications like developing the mean or plotting charts, one needs a minimum of 30 responses for answers to be valid.
- **SECONDARY DATA REVIEW:** An examination of existing data. This type of review is often the initial inquiry that precedes data collection with stakeholders. It is also called a desk review. Sources include academic theses, annual reports, independent studies by NGOs or researchers, and census data.
- **SURVEY:** A sequence of focused, targeted questions posed to stakeholders in a fixed order by a surveyor. As with questionnaires, survey questions must reflect cultural awareness and be

language sensitive in addition to fitting within a set of formal methodological standards. Surveys are generally utilized for large-scale efforts, though they may be used on a smaller scale. To reach a statistically significant conclusion, the number of responses needed depends on the total population size. However, statistical applications like developing the mean or plotting charts one require a minimum of 30 responses.

- **TESTING:** This is usually a series of questions or exercises – oral or written – for measuring the skills, knowledge, capacities, or aptitudes of an individual or group. Testing is generally used before and after training as a way to measure change.
- **PARTICIPATORY LEARNING AND ACTION TECHNIQUES (PLA)**²⁷: The application of Participatory Rural Appraisal techniques in a mutual learning process utilized on broader issues than the original rural development focus. There are many techniques within PLA that are useful for baseline, monitoring, and evaluation purposes such as:
 - **Venn diagrams:** These are made with circular cards of different sizes and colors placed in relation to one another with each card representing an issue. The size of the card represents the issue's importance to the conflict, with a larger card indicating greater importance, and the degree of overlap between cards represents the intensity of interaction of those issues. Men and women, wealthy and poor, young and old, may well produce different diagrams whose differences are often instructive.²⁸
 - **Pairwise ranking:** This technique helps to determine the relative importance of various options. The participants compare only two options at a time, and the reasons for preferring one option over the other are made clear. Going through all the possible combinations finally results in a list of criteria by which villagers can assess options.
 - **Conflict Mapping:** This is a technique used to represent a conflict graphically by placing the parties in relation both to the problem and to each other. When people with different viewpoints map their situation together, they learn about each other's experience and perceptions, and their differences and commonalities become clear.
 - **Drawing:** This technique is often called mapping as well. It is a visual depiction, generally in pictorial form, of the focus of the discussion. This can be a geographic map, an emotion, or a situation.

²⁷ Adapted from Participatory Rural Appraisal (PRA), part of the "Working instruments for Planning, Evaluation, Monitoring and Transference into Action" series (Swiss Agency for Development and Cooperation [SDC], Strategic Unit, January 1997)

²⁸ Adapted from Participatory Research for Sustainable Livelihoods: A Guide for Field Projects on Adaptive Strategies. <http://www.iisd.org/casl/CASLGuide/RepRel.htm>.

- **Role Playing:** Taking on a role enables people to creatively remove themselves from their usual roles and perspectives to portray a situation.

Since every context and application of a method can vary, it is difficult to give definitive guidance on what method to use in every situation. Some of the more common strengths and weaknesses of each method are described in the table below. Because cost is almost always a variable in selecting methods, a separate column has been added to illustrate if the method is of high, low, or average expense. Of course, this is also a question of scale: the larger the scale, the higher the cost. Thus, the cost column depicts comparative costs by assuming the same project scale is used in each method listed.

Selecting the Most Appropriate Methods

| Method | Strengths | Weaknesses | Cost |
|--------------------|---|---|-----------------|
| Direct Observation | <ul style="list-style-type: none"> • Minimal preparation required • May enable the experience of minorities or women to be captured, particularly in situations where speaking out against the norm is dangerous | <ul style="list-style-type: none"> • Must be done at the right moment in the right place • Does not provide information on why things occur • Presence of the observer may influence behaviors | Low |
| Focus Groups | <ul style="list-style-type: none"> • Can identify issues that need probing through another method • Allows one to observe various perceptions on an issue • Enables more people to be involved in less time than individual interviews | <ul style="list-style-type: none"> • Difficult to manage multiple opinions • “Group think” may occur • Individuals may not feel comfortable to dissent | Low/ Average |
| Interviews | <ul style="list-style-type: none"> • Good for small numbers • Allows for exploration into how and why • Generates data on needs, expectations, attitudes, perceptions, beliefs, and feelings | <ul style="list-style-type: none"> • Time-consuming • May be difficult to differentiate between those who are telling you what they think you wish to hear from those telling the truth | Low/ Average |

Selecting the Most Appropriate Methods

| Method | Strengths | Weaknesses | Cost |
|--|---|--|-----------------|
| Participatory Learning and Action Techniques | <ul style="list-style-type: none"> • Can offset the biases of the evaluator • Empowers the local people because their views are taken seriously • Excellent methods for working with illiterate communities • Useful if the purpose is to determine whether needs are being addressed by the evaluation (Appropriateness) | <ul style="list-style-type: none"> • No hard quantified data is produced • For comparative work replication is difficult as each situation has its own unique situation | Low/ Average |
| Photography/ Video | <ul style="list-style-type: none"> • Less open to interpretation • Easily disseminated • Can be a rapid technique • Easily led and completed by participants • Easily preserved | <ul style="list-style-type: none"> • Must be done at the right moment in the right place • Can be one-dimensional information that does not explain how or why | Low/ Average |
| Project Document Review | <ul style="list-style-type: none"> • A low-cost way of learning the history and background of a project • Provides insight into the perceptions of the practitioners | <ul style="list-style-type: none"> • May be limited in the degree of detail • May be tailored to donor or other needs and requirements but omit key information • May have gaps in time that reports do not cover | Low |
| Questionnaire (by post or e-mail) | <ul style="list-style-type: none"> • Good if the intervention affects large numbers of people • Good if statistically significant results are needed | <ul style="list-style-type: none"> • Requires literacy • Time-consuming • Is difficult to utilize in contexts with multiple languages • Requires a distribution system (e.g., postal system, Internet) for large numbers of people | High |
| Secondary data review | <ul style="list-style-type: none"> • Fast means of gathering key background information • Offers a variety of perspectives and insights • Can save the evaluation team time since they do not need to collect the data | <ul style="list-style-type: none"> • May not be tailored to the needs of the project • Data may be flawed | Low |

Selecting the Most Appropriate Methods

| Method | Strengths | Weaknesses | Cost |
|---------|--|---|------|
| Surveys | <ul style="list-style-type: none"> • Good for interventions that affect large numbers • No literacy requirements for respondents • Good if statistical comparisons are required | <ul style="list-style-type: none"> • If not well trained, surveyor bias may affect responses • Requires greater resources than a questionnaire does • Does not explain how or why something happened • Not very good if the purpose of the assessment is to study complex processes • Time-consuming | High |
| Tests | <ul style="list-style-type: none"> • Good for knowledge acquisition; less reliable for skills | <ul style="list-style-type: none"> • Does not reveal whether or not the new knowledge or skills will be retained or applied in the future • May require literacy | Low |

Data collection is based on standards of practice that are essential to the quality and, hence, the credibility of the data collected.

Data collection is based on standards of practice that are essential to the quality and, hence, the credibility of the data collected. The manner in which questions are asked and formulated, the behavior of interviewers, and identity of the interviewer (in terms of gender, nationality, race, etc.) can influence responses. The most common error made by novices is to allow their personal bias to influence the situation. The type of questions researchers ask can introduce bias as can the choice of who they talk to and when data collection is conducted. In addition, the way that data is analyzed or presented can introduce bias.

When considering how bias can affect the types and form of questions, consider, for example, a project working with French-speaking citizens of Quebec on changing their perspective from that of seeking political separation to remaining part of Canada. One of the attitudes to be changed is animosity toward the federal government. If the baseline asked, “How much do you dislike the federal government?”, there is clearly a negative bias that assumes all francophones dislike the federal government and it is simply a matter of how much. A very different response would be expected if the question were phrased as follows: “Tell me about Quebec’s relationship with the federal government.”

The manner in which questions are asked can also affect the answers given. This becomes particularly important when the evaluator and those giving information are from different cultures. Moving to the other end of Canada, the Blackfoot nation is one of the First Nations (indigenous populations) of Canada living in Alberta. If an evaluator were collecting data from the Blackfoot people, she/he would need to understand their use of silence and not rush to fill it with more questions or answers. The

evaluator who does not allow for silence and the processes of reflection within that community will return with data that is superficial at best and probably significantly flawed.

Finally, who asks the questions also influences responses. The attributes of the questioner are all noticed by the respondent and subtly interplay with the answers given. Formal education, use of special terminology, clothing, being in a paid position, being perceived to be in a position of authority, and being perceived to be sympathetic to the other side due to one's nationality are all examples of the things that may affect the responses received. Consider an older, well-dressed, male French professor going into the immigrant communities of the Paris suburbs following the November 2005 riots to gather baseline data for a project involving disaffected immigrant youth who participated in the violence. The differences between the data collector and the respondents are not fatal to the project; those differences simply need to be taken into account when designing the data collection process and designating who should collect the information.

Ways to minimize bias and errors in data collection include the careful training of researchers, setting of objectives and indicators, and the triangulation of information.

Ways to minimize bias and errors in data collection include the careful training of researchers, setting of objectives and indicators, and the triangulation of information. Because the field of social science research is well-established, there are many ways to obtain training in methods implementation ranging from university degrees in the subject to short courses and practical trainings.



How are methods selected?

Having a clear understanding of what the research exercise (e.g., baseline, monitoring or evaluation) is to explore how the resulting conclusions are to be utilized is essential to good method selection. The information required should drive the selection of methods. Use the following questions to further guide the method choice.

- **HOW COMPLEX IS THE PROJECT?** If the project is extremely complex, qualitative methods are likely to be better suited to handling its intricacies than other methods. As project complexity increases, so too does the need for triangulation.
- **WHAT IS ALREADY KNOWN ABOUT THE PROJECT?** The more data available, the more the evaluators can focus on gathering information to fill the gaps. This may be exploring the why behind a fact in which case qualitative means. On the other hand, it could be that generalized conclusions are missing; hence, quantitative means would be better suited.

The information required should drive the selection of methods.

Determining if a certain design tool was used to plan the project and whether or not the project was launched with clear assumptions and an articulated theory of change provides a sound starting point for an evaluation team. If this information is not available, time should be allotted to finding it through project document review and possibly participatory learning and action techniques with the project team.

- **WHAT INFORMATION IS NEEDED?** Does the team need information on attitudes, knowledge, perceptions, behaviors or skills? If data on knowledge and skills is required, testing might be a good approach. If behaviors are the focus, direct observation or interviews may be a good choice. If it is important for the answers to be generalized, quantitative methods are preferable.
- **HOW MUCH TIME IS AVAILABLE?** Some methods require more time to design, implement, and analyze than others. A large questionnaire, for example, takes far more time than do focus groups. A questionnaire needs to be developed, tested, redesigned, and distributed to respondents, who must be given time to fill it in. Once the questionnaires are returned, the data is generally entered into a database and then analyzed. All of these tasks not only require enough days for each to be performed, but also a sufficiently lengthy evaluation duration to allow for the data collection method to be completed. For information on time and duration estimates, see page 138 of the Evaluation Management chapter. An evaluation for which data collection must be completed in two weeks would not permit enough time for the development of a questionnaire.
- **WHAT IS THE EVALUATION APPROACH?** Although most methods can be applied within most approaches, the approach selected may lend itself better to one method over another. If the approach is self-evaluation and no one on the team has survey or questionnaire expertise, that is probably not the best option. If utilization-focused evaluation is the approach and the project team wants results that can be generalized, surveys or questionnaires would be a good method choice. However, the use of a different utilization-

focused evaluation may be desirable in order to explore why some people drop out of a program and this would best be accomplished through interviews or focus groups.

- **WHAT IS THE EVALUATION SCOPE?** The scope of the evaluation is an important factor in methods selection. If the evaluation is to cover a large number of people, questionnaires and surveys should be considered. For a medium-sized group, participatory learning and action techniques may be a good choice. If the geographic scope is large or the conclusions need to be generalized, time-intensive methods such as interviewing and direct observation would not be the best options.
- **HOW DIFFICULT WILL IT BE TO ACCESS THE DATA?** Also called data availability, the degree of difficulty involved in accessing information is always a consideration in method selection. Are the data sources out in the bush fighting a war and therefore hard to access or are they based in the city centre and accessible through the Internet? The latter may offer the opportunity to use an Internet-based questionnaire tool, while evaluation of the former would require direct observation, interviews or potentially focus groups. Do those you seek to engage speak many different languages? If so, questionnaires and surveys require translation and testing in each language, which will have cost and time implications.
- **WHEN IS THE INFORMATION NEEDED?** If there is a restricted timeframe for the evaluation, rapid methods like secondary data review, project document review or Participatory Learning and Action Techniques may be helpful. If a longer time period is available, more time-intensive methods can be used, such as surveys.
- **WHAT LEVEL OF RELIABILITY IS REQUIRED?** If a high degree of reliability is essential, Participatory Learning and Action Techniques are often not the best options, whereas questionnaires and surveys may be a better choice. The importance of method reliability increases with the emphasis on quantitative results and the ability to generalize to populations.
- **WHAT IS THE AVAILABLE BUDGET?** As outlined in the Selecting the Most Appropriate Methods table on page 207, different methods have different cost structures associated with them. Direct observation and project document review are very low in cost, while questionnaires and surveys can be expensive.
- **WHAT IS THE CAPACITY OF THE DATA COLLECTORS?** The difficulty of the method should be considered in relation to the capacity of the data collectors. If project staff members are gathering the data, selection of a simple and easily applied method is important to the quality of the data collected. Try to build on the skills

that the staff already have acquired through their project implementation work by choosing, for example, such methods as conflict mapping or photography/video. In some cases, the method may appear simple, but actually gathering data that contributes to sound decisionmaking may be more complex. The greater the degree of complexity required by the method (also called the method difficulty), the greater the level of knowledge and skills required to implement it effectively.

Triangulation is simply using different methods to research the same issue and then analyzing all of the results.

- **WOULD TRIANGULATION MAKE THE CONCLUSIONS MORE RELIABLE?** Triangulation is simply using different methods to research the same issue and then analyzing all of the results. For instance, if examining the results of a project that sought to change laws on citizenship in Ivory Coast, one could conduct in-depth interviews with members of the government, use conflict mapping with groups of non-citizens in their communities, and gather feedback forms from events. Here, the evaluation is cross-checking one result against another (i.e., triangulating), which increases the reliability of the conclusions.

Triangulation is useful in many ways. Contradictory results produced through different methods often indicate important problems with question design and/or fundamental issues surrounding the researcher's understanding of a topic. Triangulation is essential when using Participatory Action and Learning Techniques and helpful when the researcher, in exploring sensitive issues, is uncertain if the data source is able or willing to provide the full story.

The conflict and its volatility affect the choice of method.

- **ARE THERE CONTEXTUAL ISSUES SUCH AS CULTURE OR LEVEL OF DEVELOPMENT THAT WOULD MAKE SOME METHODS BETTER THAN OTHERS?** The context should always be considered in method selection. If women will not speak openly in front of men or youth cannot speak their mind in front of elders, then focus groups or surveys conducted in open environments may not be the right choice. Individual interviews or questionnaires may allow those who cannot speak their mind in all situations better opportunity to express their views.
- **ARE THERE CONFLICT ISSUES THAT MAKE SOME METHODS BETTER THAN OTHERS?** The conflict and its volatility affect the choice of method. There may be conflict situations where someone is unable to state a dissenting opinion without putting themselves in danger. In such a case, methods should be selected that ensure the anonymity of sources by not requiring the disclosure of names or documentation. Any method that requires experiences to be documented, through the use of participant diaries or photographs, for example, deserves extra consideration in conflict contexts. If discovered by the wrong people, such as a paramilitary group or the army, these participants and sources might be in danger.

Take, for instance, three selection criteria mentioned earlier – data availability, method difficulty and method reliability – and compare their implementation in stable versus volatile environments. Recall the definitions for each:

- **Data availability** reflects the ease or difficulty of obtaining data. High data availability means that it is relatively easy to access the data from the source. Low data availability would mean that is difficult to obtain the information needed.
- **Method difficulty** considers the complexity of appropriately developing and implementing the method. High method difficulty implies that there is some intricacy involved in developing and implementing the instrument, while low method difficulty means that it is relatively simple.
- **Method reliability** refers to the degree to which the method produces the same results when used by different people. High method reliability means that the instrument can be used many times and the same responses will be generated. Low reliability means that, if different people utilized the method, the answers would likely be different.

In the table on page 215, the left half of each column represents a stable environment in which a non-sensitive issue is discussed. The right half of each column represents a highly contentious conflict situation dealing with a sensitive issue.

For example, consider a project in the Ukraine that seeks to decrease negative attitudes of Ukrainians toward Russian nationals living in the Ukraine. Using focus groups would provide high data availability since accessing this information would not be difficult. In comparison, consider a similar project in Iraq in 2005 that seeks to decrease negative attitudes that Sunnis hold toward Americans. The use of focus groups would likely have low data availability because Sunni respondents would be fearful of speaking out against the accepted norm of “hatred” in front of their community.

Data Availability, Method Difficulty and Reliability²⁹

| Methodology | Data Availability | | Method Difficulty | | Method Reliability | |
|--|-------------------|-----------------|-------------------|-----------------|--------------------|-----------------|
| | <i>Stable</i> | <i>Volatile</i> | <i>Stable</i> | <i>Volatile</i> | <i>Stable</i> | <i>Volatile</i> |
| Direct Observation | High | High | Low | Low | High | High |
| Focus Group | Average | Low | Average | High | Average | Average |
| Interview | Average | Average | Average | High | Average | Average |
| Participatory Learning & Action Techniques | Average | Low | Average | Average | Low | Low |
| Photography/Video | High | High | Low | Average | High | High |
| Project Document Review | High | High | Low | Low | High | High |
| Secondary Data Review | High | Low | Low | Low | High | High |
| Survey/Questionnaires | Low | Low | High | High | High | Average |
| Tests | Average | Average | Average | Average | Average | Average |

What do I need to know about instrument development and testing?

Once the methods have been selected, the next step is to design the instrument to be used, such as the questionnaire or the interview protocol. The methodological standards for instrument development are based on a well-researched body of literature, and it is critical to carefully follow those standards. The result of not using these standards will be flawed or biased instruments that produce unreliable data. That flawed data will then create false conclusions. Since the conclusions of an evaluation are what inform program decisionmaking, the consequences of using flawed instruments can have significant negative effects on the project and the people it is meant to assist.

At this point in the process, the average practitioner should seek technical assistance if developing the instrument on her/his own or assign it as one of the tasks for the external consultant.

Once the instruments have been designed, they must be tested. Both qualitative and quantitative methods of data collection require validity and reliability tests. These check for clarity, accuracy and whether the

²⁹ Adapted from *Managing for Evaluations: A Manual for PVOs*. PVC Grants Managers and Evaluation Team Members by A. Willard (USAID/PVC/Matching Grants, 1998).

tool is appropriate to the data needs. In essence, you want to know whether or not the instrument accesses the type of data required in the way that is intended. This step is commonly omitted in peacebuilding evaluation to the detriment of the quality of results. When external consultants are hired, they should be told that they will be expected to perform validity and reliability testing and this expectation should be included in the terms of reference (see page 138 for more information on terms of reference development.) If instruments are not tested and refined, the quality of the resulting data will be compromised.

Testing can be as elaborate or as minimal as is needed by the project at hand. Assuming that your project staff members are broadly representative of both the local population and the conflict, a cheap and easy strategy is to use the project staff as subjects for the test. In other words, try the instrument on project staff, by having them fill out the questionnaire or respond to an interview. This will not only enable the evaluators to have input into refining their instrument, but it will also provide the project team with insight into what is being asked and how. Of course, if the instruments are intended to gather data from Israelis and Palestinians but the staff members of the evaluating organization are all Israeli, this would not be an effective test. Similarly, if a consensus-building project in Washington, DC worked primarily with community members who have limited educational backgrounds, testing the instruments on highly educated staff would not be appropriate. Always test the instruments in the language and culture in which they will be used.

What is disaggregated data?

In thinking about the data to be sought from the baseline, monitoring or evaluation, consider whether the conclusions would be more useful if they were broken down according to different factors such as ethnicity. These factors correspond to key groupings within the project's target population. This is called disaggregated data.

Take, for instance, a dialogue project in Northern Ireland that brings together Protestants and Catholics. Attitudinal information that represents all participants gathered through interviews will be informative, but it could hide important differences between the two communities. Potentially, if the data was broken out by Catholic participants and Protestant participants, it could reveal differences in attitudes that would be key to informing future decisions about the project.

Data can be disaggregated by many different factors or subgroups. The importance of each factor is dependent upon the focus of the intervention being explored and the evaluation objectives and lines of inquiry being chosen. Disaggregating the data within a line of inquiry should provide more useful information than if it were not disaggregated.

Once the instruments have been designed, they must be tested.

Testing can be as elaborate or as minimal as is needed by the project at hand.

Data can be disaggregated by many different factors or subgroups.

It is also important to consider the state of the conflict and the sensitivity of the factor when making these decisions.

If the information is not more useful when disaggregated, it is not worth doing it. Common factors to disaggregate against for conflict transformation projects include:

- Gender
- Residence
- Ethnicity
- Religion
- Age
- Tribe
- Ex-combatant
- IDP or Refugee

Remember that, as the number of factors increases, so too does the amount of data needed and the amount of time required to analyze it, which affects the cost of the research.

Like instrument development and implementation, data analysis is subject to strict standards of practice.

It is also important to consider the state of the conflict and the sensitivity of the factor when making these decisions. If asking someone about her/his religion or ethnicity is highly sensitive or even, in some instances, dangerous, there might be proxies that can be used to substitute for such questions as a way to avoid endangering the respondents. For instance, in some conflicts, neighborhoods are strictly segregated, so that asking where someone lives could substitute for ethnicity or tribe. The project team plays a key role in alerting external professionals to these sensitivities and to potential proxies.

Quantitative analysis can also be used on qualitative data.

Once the factors have been chosen, the appropriate questions need to be included in the research instrument in order to gather the data. For instance, if the gender and religion of participants were important to the project, these two questions would be added to a questionnaire or interview as factors to be collated. Or, if a feedback form is used to monitor a workshop, these two factors (gender and religion) would be added to the form.

How is data analysis done?

Like instrument development and implementation, data analysis is subject to strict standards of practice. Even though high-quality data collection instruments have been developed and tested appropriately, and the process of data collection may have followed all the best practices, it all will be for naught if the proper analysis techniques are not utilized.

Quantitative data analysis is usually called statistics. Generally speaking, quantitative data is processed through statistical computer software packages. The most popular of these is the Statistical Package for Social Sciences (SPSS); however, the analysis software could be as simple as an Excel spreadsheet or an Access database. With the proper safeguards for confidentiality, processing quantitative data in a public or transparent way helps build credibility in communities where people distrust the process and/or technology.

There are a number of approaches to qualitative data analysis, although none have the same precision in the rules as statistical analysis. The abundance of approaches is an asset because any set of qualitative data can be approached from a variety of different perspectives. In other words, different techniques can be applied to the same data, which may highlight new aspects of that information. Hence, the approach taken in the analysis affects the conclusions drawn.

Quantitative analysis can also be used on qualitative data. This means that qualitative data, such as transcripts of interviews, can be analyzed to produce statistical conclusions.

Identifying the means of analysis is often forgotten when discussing baselines, monitoring, and evaluation, yet it is a central part. Peacebuilding practitioners should question evaluators about their chosen means of analysis and why they believe it is the best option. For all forms of research, the author needs to be able to explain how she/he arrived at the conclusions from the data. In other words, the means of analysis must be described and scrutinized.

Analysis should not be the exclusive domain of the evaluator. The people under scrutiny in an evaluation frequently offer very insightful analysis and, at times, bring out dimensions that only they can perceive. Whoever participates in the analysis needs to be aware of her/his own biases and assumptions.

Are there unique peacebuilding tools for data collection?

In many professional fields, such as education, basic social science data collection methods are blended together or refined to meet specific needs of that field. When a standard data collection method is refined or blended with core features of a field it is called a tool. These “tools” often become the accepted techniques for baseline, monitoring, and evaluation in that field.

Few tools have been developed specifically for conflict transformation, although there are some that were developed for other fields that could be adapted. None of the tools in this latter group have taken root in the peacebuilding field, none have been deemed more or less effective than others, nor is there a norm or standard in terms of application.

Selecting a tool is not essential for an evaluation to be effective. Evaluators and project teams should consider what information is being sought and then select the best way of obtaining it. That selection may include the use of straightforward data collection methods, such as surveys or interviews, or there may be another tool that is more effective in obtaining the data.

There are a number of approaches to qualitative data analysis, although none have the same precision in the rules as statistical analysis.

When a standard data collection method is refined or blended with core features of a field it is called a tool.

Each tool needs to be adapted to the context and purpose of the intervention for which it is being used to evaluate and, as with methods, tools can be combined. Regardless of the tool used, it is necessary to have an understanding and ability to implement the data collection methods from which they are developed.

As of 2005, the Office of Transitional Initiatives (OTI/USAID) is developing a monitoring and evaluation toolkit for transition and conflict transformation projects.³⁰ A sample of the tools in this toolkit, along with several others, is listed below. Note that this is not a comprehensive list of every tool that has been developed – others are available.

ACTIVITY INTERVIEW³¹: An activity interview seeks to identify the views participants hold about an intervention’s activity, such as a Good Neighborliness Seminar, regarding the effects of that activity and/or to obtain process feedback. This input is added to the observations and opinions of the staff responsible and compiled in a short report. Using a semi-structured format, activity interviews take place a few weeks after the activity. Not all participants need to be interviewed, nor does it need to occur after each activity; rather, a sample of each may be selected. This tool would most commonly be used in monitoring.

COGNITIVE SOCIAL CAPITAL ASSESSMENT TOOL (CSCA): CSCA is “a quantitative method for collecting basic information about cognitive social capital quickly. Cognitive social capital refers to people’s perceptions of the trustworthiness of other people and key institutions that shape their lives, as well as the norms of cooperation and reciprocity that underlie attempts to work together to solve problems.”³² CSCA uses a questionnaire, which can also be utilized as a survey. It can be implemented in small- to large-scale applications. The quantitative nature of the tool may fail to capture some of the complexities of social change. This tool is most appropriate for a baseline and evaluation.

MEDIA CONTENT ANALYSIS TOOL: This tool allows project managers to “evaluate media coverage, placement of stories, tone, and visual images, prominence of quotes/personalization, and reach of a media outlet. It can be used to track how different media cover topics such as conflict, human rights, reintegration of ex-combatants, and local governance reform.”³³

To utilize this tool, the evaluator identifies the media outlets to include, determines their circulation/reach, and then scores each against seven measures: prominence, headline, visuals, quotes, tone, column inches, and political ideology. The score for each outlet is then multiplied by

³⁰ At the time of this manual’s publication, the *OTI Performance Monitoring and Evaluation Toolkit, Fast Learning for Program Improvement, Local Empowerment and Results* [hereinafter *OTI Toolkit*], is not yet available publicly. Inquiries should be made to Social Impact at <http://www.socialimpact.com/>.

³¹ Adapted from *Strategic and Responsive Evaluation of Peacebuilding: Towards a Learning Model, Report of the Second Action-Reflection Seminar Convened by NPI-Africa and the NCCK-CPBD Project*, March 2001.

³² *OTI Toolkit*, forthcoming, pp. 146.

³³ *OTI Toolkit*, forthcoming, pp. 111.

the outlet's individual ranking in terms of circulation/reach. The scoring measures may need to be adapted to those issues of importance to conflict transformation. This tool is best used for baseline and evaluation studies, though it could be modified to contribute to monitoring efforts.

CASE STUDY: This tool investigates a contemporary event within its real-life context. Case studies are a way to learn from past experience since they explore how something happened. A case study results in a report that contains a rich narrative of the phenomenon detailing how it came about.³⁴ It is based on a particular worksite the boundaries of which need to be clearly defined to allow the study to be focused. If the case is to be illustrative of the wider context, selection of a site that is not unique is important.

Case studies rely on multiple data sources because the data needs to converge in a triangulating fashion. A case study involves the use of a variety of data collection methods, predominately interviews and project document review, but can also include direct observation or focus groups. Case studies can be done on a rolling basis to monitor the changes that occur over time or as part of an evaluation both of which result in brief, reflective snapshots of complex and dynamic situations.

CAPACITY ENHANCEMENT NEEDS ASSESSMENT (CENA): CENA "is a participatory assessment [tool] designed to evaluate existing capacity within key community and local government stakeholder groups, identify capacity gaps and weaknesses and recommend possible remedies."³⁵ Based on interviews and focus groups, information from a CENA is then plotted against each indicator on a scorecard. The tool was created for community-based development and, as such, the indicators would need to be adapted for conflict transformation programming. This tool would be best used in baseline and evaluation studies, though a streamlined version might be possible for monitoring efforts.

FOUR LEVELS OF TRAINING EVALUATION³⁶: The four levels approach - reaction, learning, transfer and results - is a systematic way to assess the quality and outcome/impact of training. Information from the previous level serves as a base for the next level as one works through all four. Reactions (level one) should never be the only level utilized and can often be blended with learning (level two). Generally, reaction (level one) utilizes a questionnaire and learning (level two) involves a pre- and post-test, while transfer (level three) can either use interviews or surveys several months after the training has taken place. Levels one and two should be included in monitoring systems while levels three and four might be a monitoring or evaluation tool.

³⁴ Robert Yin, *Case Study Research: Design and Methods*, Third Edition, Thousand Oaks, Sage Publication, 2003.

³⁵ *OTI Toolkit*, forthcoming, pp. 55.

³⁶ Adapted from the *OTI Toolkit*, forthcoming.

Can I draw a conclusion for the entire population from this data?

How to draw conclusions for an entire population, be it a community, tribe, or geographic area, is beyond the scope of this introductory manual. Conclusions are drawn using quantitative methods and should be based upon “enough” of a representative or random sample of the population. What constitutes “enough” requires a calculation based on established practices.

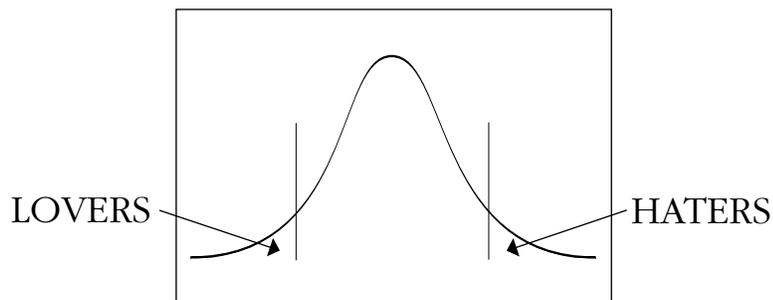
In other words, they are not representative of the average – they are lovers or haters.

As practitioners who will be gathering data through monitoring practices, however, there is one related concept that is important to understand. It can be called informally *Lovers, Haters, and Everybody Else*. As depicted below, the bell curve represents the average population of people, with the majority located in the middle and the extremes located at either end. Extreme means not being representative of the average or norm within the group. Although it appears that the extremes are the same size at either end, this is not necessarily the case.

Lovers, Haters, and Everybody Else relates to conflict transformation monitoring because it explains the implications of basing our monitoring data collection only on participant-initiated measures such as personal narratives, testimonials, fan mail, individuals who call into radio shows, or informal comments from participants. As a general rule of thumb, individuals who are willing to step forward and to take their own time to volunteer information fall at either end of the bell curve. They either love what is being done and want to praise it so it continues or they hate what is happening and feel compelled to intervene to stop it. In other words, they are not representative of the average – they are lovers or haters.

Classifying such individuals as lovers or haters does not discount their input. What it does tell us is what those two ends of the spectrum think about the work. We cannot take that work to the next step and assume that our information is representative of everyone involved.

The Lovers and Haters and Everybody Else



What record maintenance systems are necessary for collected data?

Record maintenance for baseline, monitoring, and evaluation data is quite straight-forward. For baselines and evaluations, the raw data collected, such as completed questionnaires or interview notes, should be kept in its hard-copy form. Future evaluators and researchers may wish to go back to the original data and do a different analysis or verify the previous one. If the data is sensitive or confidential, it should be stored appropriately, either in a locked filing cabinet or an inaccessible room. How long raw data is kept is dependent upon the project, organization, and potential future uses. If a formative evaluation was performed and a summative evaluation will occur later, the records should be saved until the summative evaluation is done.

Evaluators often do not return raw data to the project team unless it is requested. Particularly in the case of baseline and formative evaluation, retaining the raw data should be part of an organization's good practice norms.

For the material itself, one can use well-labeled filing cabinets, computer floppy disks, or CD-ROMs to preserve the information. Generally, it is good practice to keep a back up file of all electronic documents. If the data was processed using Excel spreadsheets or SPSS, keep those electronic files because it will save data entry time for future efforts.

When considering record maintenance systems for monitoring data, simple is always better. If most project staff members are comfortable with using an Excel spreadsheet, select that as the medium for data analysis rather than a statistical package that requires a specialist to operate.

What are the ethical obligations of feeding back the results to the people involved?

When people hear nothing about a study or evaluation in which they participated, they are less inclined to contribute to future data collection efforts. This is even more the case in highly participatory exercises where people feel they have an investment in the outcome of the research. Even the briefest feedback on the general findings and the use of those findings is often appreciated very much.

Future evaluators and researchers may wish to go back to the original data and do a different analysis or verify the previous one.

If the data is sensitive or confidential, it should be stored appropriately, either in a locked filing cabinet or an inaccessible room.

Further Reading:

Chronic Poverty Research Centre Methods Toolbox

<http://www.chronicpoverty.org/CPToolbox/toolboxcontents.htm>

William M. Trochim, *The Research Methods Knowledge Base*, Atomic Dog Publishing, 2001. <http://www.socialresearchmethods.net/kb/>

Keith Punch, *Introduction to Social Research Quantitative and Qualitative Approaches*, SAGE Publications, 2000.

CONCLUSION

“It is not that we should simply seek new and better ways for managing society, the economy and the world. The point is that we should fundamentally change the way we behave.”

- VACLAV HAVEL

An historian once said, “Most of our so-called reasoning consists in finding arguments for going on believing as we already do.”³⁵ Monitoring and evaluation offer the opportunity to counter this tendency, to open ourselves to new information, and to engage in responsible analysis. To exploit this opportunity, greater effort must be made to incorporate design, monitoring, and evaluation into conflict transformation programs. This effort will serve the peacebuilding field by increasing our understanding of, and ability to prove change in complex conflict situations.

Much has been learned from DM&E in peacebuilding so far...
.....

When Boston wanted to stop youth violence and homicide, a partnership – composed of researchers, community leaders, members of the clergy, probation officers, police officials, and federal enforcement agency personnel – came together to devise a strategy to intervene in the local gun market. When data revealed that the problem was more specifically caused by youth gangs, not simply gun markets, the partnership adjusted its strategies. Boston’s hard work paid off: youth homicides fell by two-thirds after the ceasefire strategy was put in place.³⁶
.....

In considering the importance of involving local authorities in refugee reintegration projects in Rwanda and Bosnia, an evaluation team concluded: Local and regional authorities were taken into account in different ways. Some were included, some excluded intentionally, and some ignored. Where it was possible to include them constructively, they became important allies in promoting coexistence. Where they were ignored, they undermined success.³⁷
.....

More dedicated thinking and testing of DM&E techniques is necessary so that they, too, become increasingly effective for peacebuilding. Though strides have been made, there is still much to learn in order for DM&E to reach its full potential as a learning tool within conflict transformation.

One can learn to build with stone by reading books and experimenting. The principles of laying stone are few and easy to follow – enough

³⁵ James Harvey Robinson, *The Mind in the Making*, http://en.thinkexist.com/quotes/james_harvey_robinson/.

³⁶ National Institute of Justice, *1998 Annual Report to Congress*, December 1999.

³⁷ Eileen F. Babbitt et al., *Imagine Coexistence: Assessing Refugee Reintegration Efforts in Divided Communities*, The Fletcher School of Law and Diplomacy, Tufts University, July 2002.

so that progress comes quickly in making walls both aesthetically appealing and structurally sound. By taking apart a stone house built by experienced masons, the novice stone layer discovers how the veterans dealt with more nuanced and challenging issues. The novice can then recycle those stones and build another house in ways she/he had not known previously.

We've tried to disassemble design, monitoring, and evaluation for peacebuilding by taking each subject apart "stone by stone." Every reader will have to gather the stones presented here, and others from elsewhere, to build their own structure. Hopefully, the construction will include approaches and ideas that our readers had not known or previously practiced.

Of course, many of the building blocks needed to succeed in peacebuilding have not been addressed in this manual. Our intent is not to tell anyone how to transform conflicts, but rather to illustrate the range of choices in peacebuilding that catalyze needed change and the techniques available for that work to contribute to learning within the field.

We've maintained throughout this manual that monitoring and evaluation are the most accessible learning disciplines available to peacebuilding. In putting this manual together, we're reminded of another valuable learning discipline: writing. We've forced ourselves to support our beliefs with examples, to check our jargon, and to think backwards and forwards on how we've done and will do many of the tasks here described. We've done less well at shedding our Western paradigms and values. The acts of writing, explaining, and connecting thought and action have proven to be very educational for the authors, as they can be for all peacebuilders.

In closing, we would like to draw attention one more time to a number of themes that run throughout this manual:

- Peacebuilding and DM&E are all about change.
- The most brilliant and creative ideas only become the best alternatives to bringing about change when they incorporate sound practices in design, monitoring, and evaluation.
- Peacebuilding is a unifying process. Successful design, monitoring, and evaluation bring together the parties, practitioners, designers, and evaluators. The artists and the technicians together can produce quality peacebuilding initiatives.
- We can all make better decisions. The data collected through a baseline, monitoring exercise, or an evaluation should inform our decisions, and better decisionmaking can, in turn, improve our work.

- Knowing when to get help is important. Knowing how to use help is imperative.
- DM&E is an opportunity. Baselines, monitoring, and evaluation provide the peacebuilding field with a valuable opportunity to show policymakers, the public, the press, and the people who we work for – the stakeholders – that conflict transformation produces positive results.

In closing, the old adage for university professors, “publish or perish,” comes to mind. For peacebuilding, there is a new standard rightly being imposed: demonstrate effectiveness or perish. The concepts, tools, and examples in this introductory manual are intended to help peace workers of all stripes rally to this call for results. We believe peacebuilding works and that we have a responsibility to show others the results.

“This is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.”

- WINSTON CHURCHILL

APPENDIX A

SOURCES FOR THE TERMINOLOGY DECODER

| Organization | Source | Website |
|---------------------|---|---|
| SFCG | | http://www.sfcg.org |
| UNDP | UNDP Programming Manual, January 2003 | http://www.undp.org/bdp/pm/chapters/progm4.pdf |
| CIDA | CIDA The Logical Framework: Making it Results-Oriented December 2002 | http://www.acdi-cida.gc.ca/cida_ind.nsf/0/c36ebd571b6fa02985256c620066cd6f?OpenDocument |
| EuropeAid | EuropeAid Aid Delivery Methods: Project Approach “The Logical Framework Approach” March 2004 | http://europa.eu.int/comm/europeaid/qsm/project_en.htm#2.%20Logical%20Framework%20Approach |
| TACIS/ EuropeAid | EuropeAid Standard Call for Proposal Documents “Annex C: The Logical Framework” 2003 | http://europa.eu.int/comm/europeaid/tender/gestion/pg/e03_en.htm |
| DFID | DFID Tools for Development Section 5.3 September 2002 | http://www.dfid.gov.uk/pubs/files/toolsfordevelopment.pdf |
| World Bank | World Bank Log Frame Handbook January 2005 | http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2005/06/07/000160016_20050607122225/Rendered/PDF/31240b0Lfhandbook.pdf |
| UNHCR | UNHCR Project Planning in UNHCR: Practical Guide March 2001 | http://www.the-ecentre.net/resources/e_library/doc/Project%20Planning%20in%20UNHCR.pdf |
| SIDA | SIDA The Logical Framework Approach 2004 | http://www.sida.se/Sida/jsp/polopoly.jsp?d=1265&a=16274 |
| RELEX | RELEX ECHO Manual Project Cycle Management October 2003 | http://europa.eu.int/comm/echo/pdf_files/partnership/pcm_echo_en.pdf |
| SADC | Swiss Agency for Development and Cooperation Setting It Up: Definition Logical Planning Framework (no date) | http://web.mit.edu/urbanupgrading/upgrading/issues-tools/tools/Log-Plan-Framework.html |
| EIDHR | European Initiative for Democracy and Human Rights | http://europa.eu.int/comm/europeaid/projects/eidhr/cfp-micro_en.htm |
| Danida | Danida Logical Framework Approach: A Flexible Tool for Participatory Development 1996 | http://amg.um.dk/en/menu/TechnicalGuidelines/LogicalFrameworkApproach/ |
| USAID | Building a Results Framework, USAID Center for Development Information and Evaluation | http://pdf.dec.org/pdf_docs/pnaca947.pdf |

APPENDIX B

EVALUATION TERMS OF REFERENCE (TOR) CIRCULATION OPTIONS

Evaluation field

1. M & E News <http://www.mande.co.uk/cgi-bin/www.mande.co.uk/forum.pl>
2. PARC (Performance Assessment Resource Center): Submit inquiry through <http://www.parcinfo.org/enquiry.asp?subject=consultancy>
3. International and Cross Cultural Evaluation Topical Interest Group of the American Evaluation Association: XCeval@topica.com
4. AEA Job Bank <http://www.eval.org/JobBank/jobbank.htm>
5. African Evaluation Association: <http://www.afrea.org/>
6. The Evaluation Center EVALJOBS listserv <http://evaluation.wmich.edu/archives/index.html>
7. Brazilian Evaluation Network <http://www.avaliabrasil.org.br/IntroductionRebramaEnglish.html>
8. Canadian Evaluation Society listserv: <http://www.evaluationcanada.ca/site.cgi?s=4&ss=6&lang=en>
9. UK Evaluation Society <http://www.evaluation.org.uk/>
10. European Evaluation Society <http://www.europeanevaluation.org/>
11. Malaysian Evaluation Society <http://www.mes.org.my/>
12. Nigeria Monitoring and Evaluation Network <http://66.201.99.156/afrea/content/index.cfm?navID=3&itemID=735>

More national evaluation societies, which do not have listserv or job board, can be found at <http://www.mande.co.uk/societies.htm>

Conflict Transformation field

- Alliance for Conflict Transformation

<http://conflicttransformation.org/Home/SubscribeTermsofService/PostYourAnnouncement/tabid/175/Default.aspx>

- Alliance for Peacebuilding: <http://www.aicpr.org/>

Development field

- Relief Web <http://www.reliefweb.int/vacancies/>
- DevNetJobs. <http://www.devnetjobs.org/>
- ConsultingBase: <http://www.consultingbase.com/>
- Development Opportunities <http://www.dev-zone.org/jobs/>