

The background of the top half of the cover is an abstract, textured composition. It features a central globe-like shape with a grid pattern, overlaid with golden, ornate arches and patterns that resemble traditional Islamic or Middle Eastern architecture. The color palette is dominated by various shades of teal, blue, and green, with golden highlights and some reddish-orange accents. The overall effect is one of complexity and global interconnectedness.

TRANSFORMATIONAL EVALUATION

FOR THE GLOBAL CRISES OF OUR TIMES

Rob D. van den Berg

Cristina Magro

Marie-Hélène Adrien

EDITORS



IDEAS

KNOWLEDGE CAPACITY DEVELOPMENT NETWORKING

TRANSFORMATIONAL EVALUATION

FOR THE GLOBAL CRISES OF OUR TIMES

EDITORS

Rob D. van den Berg

Visiting Professor, King's College London
Leidschendam, the Netherlands

Cristina Magro

International Evaluation Academy, Member of the Council
Belo Horizonte, MG, Brazil

Marie-Hélène Adrien

Universal Management Group, Senior Associate Consultant
Montreal, Canada



© 2021 International Development Evaluation Association (IDEAS), Exeter, UK
Email: ideascoordinator@gmail.com

All rights reserved.

Any views expressed in this book are those of the authors. They do not necessarily represent the views of IDEAS, the editors, the authors' institutions, financial sponsors or reviewers.

This book is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY NC-ND 4.0): <https://creativecommons.org/licenses/by/4.0/legalcode>

This license permits any non-commercial use, duplication, adaptation, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source. Please cite the work as follows:

Van den Berg, Rob D., Cristina Magro and Marie-Hélène Adrien (eds.). 2021. *Transformational Evaluation for the Global Crises of Our Times*. Exeter, UK: IDEAS.

ISBN (paper): 978-1-9168982-0-2

ISBN (electronic): 978-1-9999329-9-2

Assistant editor: Zuzana Vozárová

Copy-editing: Ann Shildneck

Design: Nita Congress



CHAPTER 12

Multiple Dimensions of Evaluation in Fragility, Conflict and Violence

INGA-LILL ARONSSON AND HUR HASSNAIN

Abstract. Evaluation in contexts affected by conflict and fragility is political and complex and can exacerbate violence. In such unpredictable environments, understanding how change happens is challenging because different actors at local, national and international levels have varied interests and definitions of what change is or may be. Reports on progress on the Millennium Development Goals (MDGs) show that these contexts lag in establishing robust monitoring and evaluation systems. Most examples of tools and resources available are from 'normal' contexts and do not fully support identification of the multiple biases at all levels during an evaluation exercise, including possible bias of the evaluator. In these contexts, collaboration with people on the ground is paramount to contextualize scenarios, tools and values in order to visualize what works or not for evaluation in a particular setting. The authors combine in this chapter background research completed, a reflection of their own personal experience, and the rich discussions with global evaluation practitioners who have worked in some of the worst conflict-affected contexts. During the IDEAS's Global Assembly in Prague, the authors conducted a one-day workshop on evaluation in fragile, conflict and violent contexts was attended by participants from many countries representing a range of leading actors and organizations that enriched this urgent topic.

Introduction

In this chapter, we critically reflect on some of the multiple dimensions of evaluations in fragility, conflict and violence (FCV) and in situations of global pandemic such as COVID-19, to *problematize* evaluation in environments that are fluid, complex, unpredictable or violent, with the aim of achieving transformational change within the broader frame of the Sustainable Development Goals (SDGs). Evaluation for transformational change is no easy task. As we argue here, it is even more difficult in FCV contexts, as well as during unstable times or crisis. In these circumstances, people are vulnerable; lack trust; live in complicated and sensitive relationships and lack faith in experts, outsiders, locals and government representatives alike. In a worst-case scenario, interaction with the wrong person could put one in danger or at minimum make it impossible to gather and process critical information and act on it, as the current pandemic has shown.

Hope, too, needs to be provided, and the Organisation for Economic Co-operation and Development (OECD) inspires it by commending the 2030 Agenda:

The Sustainable Development Goals set a roadmap for a better world. One where poverty, hunger, disease, climate change and gender inequality are no longer a threat to our planet and wellbeing. Instead, they chart a world where decent jobs for all, sustainable infrastructure, clean oceans and energy, responsible consumption and production, clean water and sanitation, and quality education, become the norm (OECD 2019, 3).

Most of us certainly want that better world that Agenda 2030 proposes. Nevertheless, from an evaluation perspective within FCV situations, the above quotation is problematic – because of its limitless optimism. This message would benefit from a touch of pessimism. From a realistic perspective, the SDGs might be difficult to achieve, partly because of the pandemic; Oxfam (2020) estimates that the crisis could push half a billion people back into poverty unless urgent action is taken to bail out developing countries. There is a risk that people may turn against the SDGs if they are perceived as unrealistic, or even belittling towards ordinary people, especially when considering the time frame of 2030. It may not be necessary to call the SDGs ‘worthless’, as Easterly (2015) did and as discussed in Van den Berg, Magro and Salinas Mulder (2019a), but there is a gap between the ‘real world’ and ‘real world evaluations’. This chapter aims at contributing with reflections on evaluation as a possible maker of transformational change, towards achieving the SDGs in the most difficult and confounding

circumstances, such as situations of FCV. The recently launched Global Evaluation Initiative, led by the World Bank and the United Nations Development Programme, estimates that only one-third of the countries who committed to achieving the SDGs have the monitoring and evaluation skills and capacities to do so (GEI 2020).

Evaluation in FCV is mostly – but not exclusively – associated with SDG 16, on peace, justice and strong institutions. Evaluation in FCV with respect to the SDGs must address complexity and unpredictability which are challenging in practice and theory in the evaluation field. This is not to claim that complexity and unpredictability are absent in normal circumstances, but in FCV contexts, specifically when transformations for the SDGs are a concern, this complexity and unpredictability are intensified, tighten and change quickly, and the stakes are high for all involved. Evaluators must get it reasonably right.

We began joint work on evaluation in FCV a few years ago, when we met at the 2017 IDEAS Global Assembly and realized that our different professional and geographical backgrounds offered us a fruitful platform for intellectual exchange and collaboration (see e.g. Aronsson and Hassnain 2019; forthcoming a; forthcoming b). This chapter will not offer a literature review on evaluation in FCV but rather will build on our experiences and the experiences of others with whom we had the opportunity to exchange ideas and perceptions in various evaluation fora using cases from the Global North and the Global South.

To lay the ground for critical reflections, we will begin by presenting a brief frame for FCV evaluation. We will then indicate challenges we faced evaluating in FCV and formulate a handful of guiding statements that summarize our first-hand experience. These were widely discussed and validated at the workshop that two of us facilitated at Evaluation for Transformative Change: Bringing Experiences of the Global South to the Global North, held in Prague in 2019¹. This conference was a joint effort of the IDEAS Global Assembly and the Third International Conference on Evaluating environment and Development.

Furthermore, by presenting a case of evaluation on violent extremism in Europe and a case on COVID-19 in Europe, we indicate that there are overlapping and distinct analytical levels between stable and unstable contexts that are worth exploring because we have not identified all the

¹ Multiple Dimensions of Evaluation in Fragility, Conflict, and Violence (EvalFCV). Workshop at IDEAS Global Assembly. Conference on Evaluation for Transformative Change. Prague, 1 October 2019.

mechanisms to be considered in FCV evaluations. Experts will probably not identify *all* factors involved, and some imperatives may vary from one situation to another. Nevertheless, a handful of precise indications are necessary and possible, as our experience in evaluation in FCV contexts and crisis show.

Collaboration is a key message in Van den Berg, Magro and Salinas Mulder's (2019b) book on transformational change. It is argued that collaboration is needed not only in the participatory methodology on the ground, but also between academic fields and evaluation units and between practitioners and theorists to make high-quality, viable progress towards the SDGs. At the same time, silos of knowledge on particular topics are essential building blocks of scientifically grounded work, which means that a required systems perspective and effective collaboration between experts from different areas of knowledge have not penetrated the evaluation field to a level that achievement of the SDGs require.

A Frame for Evaluation in FCV

It is estimated that 20 per cent of the world's poor are living in regions affected by FCV and that, by 2030, at least 46 per cent of the world's poor will be living in such regions (World Bank 2017) or more than 80 per cent if action is not taken (OECD 2018). These estimations are from before the COVID-19 crisis. Considering the many predictions being made, for example the Oxfam estimates discussed above, it seems likely that this pandemic will have dramatic effects on people living in countries affected by FCV and will push some additional states towards instability and fragility.

The SDG-16 Progress Report (Institute for Economics and Peace 2017) highlights that FCV-affected countries were on average 25 per cent more likely to have missed their Millennium Development Goals (MDGs) than other countries. It further explains that:

MDG indicators for which the majority of fragile and conflict-affected countries recorded the poorest results were those that addressed child mortality, maternal health and environmental sustainability. No conflict-affected country achieved the goal of reducing by two-thirds the under-five mortality rate between 1990 and 2015. Additionally, many of the fragile and conflict-affected countries have difficulty in maintaining the necessary systems to adequately capture the data. This can lead to poor quality data, resulting in situations appearing worse or better than what they are (Institute for Economics and Peace 2017, 7).

In evaluating the Aiding the Peace Initiative, Bennet et al. (2010) report that, in South Sudan, the support that multiple donors provided during 2005 to 2010 was often mistargeted. Reasons for this could be that donors did not fully take into account the key drivers of violence; essential services were overemphasized and security, policing and the rule of law were relatively neglected. With so many people living in vulnerable conditions worldwide, it is crucial that the right conditions be created to ensure rigorous, sensitive data collection and evaluation in FCV contexts, especially concerning achievement of the SDGs and to ensure informed decision-making in global pandemics (e.g. COVID-19).

Achievement of the SDGs in FCV states calls for a comprehensive repository of tools and resources to facilitate learning about different approaches to evaluating in FCV while making these evaluations climate and gender sensitive. Evaluation challenges in these contexts include, but are not limited to, difficulties in identifying and accessing affected populations because of rapidly changing context or moving populations; understanding power and relationship dynamics; facing fear and sensitivity regarding fact-finding missions and felt grievances; limited availability of good-quality data; paying attention to unintended effects; lack of appropriate tools and resources; and signs of corruption and human rights violations that are difficult to validate and report on. Other problems may include identifying competent evaluators who are willing to travel to conflict-affected areas and not being able to maintain impartiality throughout the evaluation process given the political context and difficulty in engaging with all key players (Hassnain et al. 2021). In addition, establishing indicators and their targets and measuring them are extremely challenging, for the many reasons discussed above. Peace, for example, is inherently political. There are multiple definitions of and perspectives on what peace is; it is perceived differently within and between countries and regions (Hassnain 2017). Furthermore, understanding and measuring the difference between targeted results and overall achievements in environments with FCV can be challenging. First, as discussed above, these states in FCV situations do not have sufficient resources for monitoring and evaluation, but even if they did, performance tracking systems that are mainly designed to measure the results of assistance in normal circumstances could easily miss progress – ‘Sometimes just keeping the lights on can be considered as a success’ in FCV environments (Kelly, Nogueira-Budny and Chelsky 2020). These issues are further described below.

Key Challenges of Evaluation

Programmes in contexts of FCV operate in a sensitive political environment and address complex challenges that are often hard to measure and report on. To evaluate them, the first task evaluators have is to build a comprehensive, systemic understanding of the context in which they are to work, including underlying cultural, social, economic and political factors and their interplay (Hassnain 2017; Aronsson and Hassnain 2019), but it does not end there. We have identified a variety of immediate challenges that evaluators must address:

- *Defining how change happens.* Evaluation in FCV contexts has weaker theoretical foundations and a limited evidence base of what works and what does not than in normal contexts. In such situations, programmes are founded on assumptions based on the principle that, if things work well, repeat, and otherwise drop them.
- *Fluidity.* FCV contexts are unpredictable, transitional in nature and fluid; hence it is difficult to submit them to a comparative analysis. Hassnain reports on his experience in Afghanistan in 2016, when many Afghans began to return from neighbouring countries. Many did not have a clear plan as to where to go to settle. Baseline data were gathered where they were initially located, but most moved to more convenient locations, such as where they had relatives or affiliations, before the intervention began. A similar situation occurred in South Sudan in 2017. These movements of beneficiary groups may mean shifts in operational priorities, resulting in limited or no reporting or sharing of such experiences while working on the ground. Aronsson reports on her experiences from the Zimapán resettlement project in Mexico, where counting affected people was difficult because they are constantly moving. Regardless of thoroughness, there always seemed to be an error at the end.
- *Risk of violence.* The imminence of bursts of violence limits face-to-face meetings for interviews and focus group discussions or even conducting surveys. Hassnain describes an experience in the Swat region of Pakistan in 2009, where a focus group discussion in a Taliban-affected community needed to be stopped because of the fear of escalating violence between the participants. Aronsson describes a violent uprising with gender dimensions in a resettlement project in central Mexico (Aronsson and Hassnain forthcoming a)

- *Trust.* Evaluations of FCV contexts and of peace, justice and strong institutions are fundamentally political and therefore fundamentally contested. Under different political perspectives, different things would be chosen to monitor and measure. Under those fluid, unpredictable circumstances, to reach agreement on indicators, tools and contextualized data is difficult and time consuming. In FCV contexts, everyone has their own agenda, and no one is neutral. To derive conclusions on the basis of sufficient triangulation efforts, as the best evaluation guides recommend, becomes of ultimate importance.
- *Learning.* What works today in a fluid context may not work tomorrow. In addition, there is limited or no culture of learning and sharing in such contexts because the contexts are changing rapidly and because of shifting priorities, low literacy levels, lack of trust between different parties and challenges related to data confidentiality. Documenting and reporting in such situations can be sensitive because of the politicization of international involvement and political sensitivities in national contexts, and evaluators may find it difficult to maintain a safe, credible evaluation space for learning and sharing.

The Guiding Statements

The following guiding statements provide a useful path for evaluators about to face a FCV situation. Although many of these statements may be valid in any evaluation, FCV contexts require that extreme attention be paid to each of them.

1. *The evaluand and the evaluator together determine the evaluation methods and approaches.* Do not get carried away by glittery tools and methods or smooth talkers.
2. *The context defines the methods.* Programmes in contexts of FCV operate in a political environment and address complex issues that are often hard to measure and report on. To evaluate them, the first task of evaluators is to have a comprehensive, holistic understanding of the context in which they work, including underlying cultural, social, economic and political factors and their interplay.
3. *People always come first.* The safety of respondents – individuals and institutions – evaluators, including enumerators, takes precedence over any kind of accountability measures. Consider seriously,

in advance, whether it is feasible and safe to conduct an evaluation in the given context.

4. *Be realistic and try to balance sensitive and objective knowledge and norms.* This is valid for both local and expert knowledge.
5. *Never reduce evaluation and social relationships in FCV contexts to a pedagogic or learning experience.* Follow an engaging, participatory process that makes a difference.
6. *The evaluators and enumerators must have documented ethical protocols, integrity and specific capacity and skills to work in FCV contexts.*
7. *Adaptive management approaches lead to iterative learning and context-responsive adjustments.*
8. *Participatory monitoring at regular intervals keeps the evaluators mostly informed of real-time changes on the ground, helping make sense of data over a given timescale.* Embedded evaluation or monitoring as a project activity, for which implementers are accountable and that is linked to feedback loops, may help the evaluator obtain data and make sense of it.
9. *Monitor for key factors identified through conflict analysis (e.g. dividers and connectors in conflict situations), but be prepared for participatory monitoring to elicit information on additional social impact issues, including gender and humanitarian protection.*
10. *Use information and communications technologies, mixed-methods and goal-free approaches in FCV contexts and situations of pandemic in which the relationship between the causes and effects is complex (Hassnain and Lorenzoni 2020).*

IDEAS's *Evaluation in Contexts of Fragility, Conflict and Violence: Guidance from Global Evaluation Practitioners* (Hassnain, Kelly and Somma 2021) discusses these challenges in detail and asks important questions such as 'Who are the potential beneficiaries of the evaluation?' 'What are the unintended consequences/effects?' and 'What can be learned?' These questions may be addressed using mixed-methods, goal-free approaches such as *outcome harvesting*, working backward to the intervention logic and looking for *how* and *what* the intervention contributed. Moreover, using information and communications technology to prevent direct contact with people, where possible and feasible, is extremely helpful.

We agree with Chigas et al. (2006) that *trust* is a concept that could guide us in evaluation in fragility and conflict. Trust is universal, because without it, no social group could collaborate, which is part of the reason for

the success of humankind. Nevertheless, it is a morally and culturally constructed concept that could be misunderstood – or intentionally misused to manipulate people. In both cases, the evaluation turns into a pedagogic learning exercise more for the benefit of the evaluation team and the principals than for the intended benefactors. The evaluation commissioners and evaluators must find a fair balance by engaging as many local people as possible not only in collecting data in difficult situations, but also throughout the evaluation cycle, from study design to analysis and dissemination. This will not only make sure that the findings are relevant, but will also build local evaluation capacities.

Few professionals and academics would disagree with the above-presented dimensions of FCV evaluation. Its strength is the compilation and insights based on our experiences and on conversations with evaluators with or without experience in FCV contexts, but these identified dimensions of FCV must be further validated, especially concerning their use as instruments for transformational change.

A Pedagogy for FCV Evaluation Training

The participatory workshop we facilitated at the 2019 IDEAS Global Assembly gathered 25 evaluation experts from around the world, to whom we are indebted for the generous contribution through the exchange of lived experiences, ideas and knowledge. They represented bi- and multi-lateral agencies, including the United Nations, the World Bank, civil society organizations, think tanks and academia, from 23 countries, including some of the worst conflict-affected countries in the world, such as Afghanistan, the Occupied Territories of Palestine and Pakistan.

The workshop allowed us to take a step forward in our inquiry into evaluation in FCV states, which was a critical examination of the pedagogy for FCV evaluation. We explored insights, methods and theoretical implications for evaluation in FCV environments that can also be applied in situations such as global pandemics, where access to respondents is almost impossible.

Although one of our main objectives in facilitating this workshop was to introduce learning tools for enhancing practice for evaluation in FCV, relevant theoretical insights were also explored. The participants were eager to understand what, why and how things mattered in FCV in a different way than in other evaluation settings. Rather early in the workshop, participants and facilitators agreed that evaluation methods in FCV must be 'real

world evaluations' (Bamberger 2007) to adequately advise policymakers and practitioners on what works in the changing, complex contexts of FCV and, more importantly, how to avoid expensive mistakes. The challenges and guidelines presented in the previous sections were discussed and validated. Thereafter, a toolbox for evaluation in FCV was established in a spirit of joint understanding about participatory consultation, acceptance and agreement. This means that the tools are available, were shared with a relatively wide audience and can be used. Still, there is this nagging feeling that a roadmap must still be completed.

An implementation can fail for many reasons, but there is an interesting discontinuity between practitioners' practice on the ground and what they know and how the results are presented in the evaluation reports (especially the published versions). In the workshop, we found a wealth of tacit knowledge that workshop participants displayed in the dialogues but seemed to avoid articulating in the reports.

We also discussed the hesitation to raise, for example in focus group and individual interviews, sociopolitically sensitive topics that might be regarded as politically incorrect or were seen as a no-go zone for discussion. The idea of transformational change means that any cultural misguided sensitivity must be scrutinized and eliminated. Would this mean that the evaluator talks from a power position? To the contrary, the authors of this chapter reason that, to reach any transformational change, it is necessary to talk about difficult things. It is condescending not to discuss with the people involved, but such discussions demand considerable sociocultural competence and personal integrity. It would be a misjudgement to claim that all local evaluators possess the necessary social and cultural skills, as likewise it would be to claim that all external experts have competence, or for that matter, that the local affected people have it. Communication and collaboration are tools in a global interconnected world, but discussions must be honest and fearless; accountability, obligations and rights go hand in hand.

Based on the above, it can be argued that the FCV evaluation toolbox might lack aspects of relevant methods of FCV evaluations or that we are chasing something that can never be caught and presented in evaluation reports. There will always be a backstage and a frontstage. Here, theory enters the stage, because having a toolbox without adequate understanding of reality will only make us repeat ourselves in complicated hermeneutical circles or end up in power and identity politics cul-de-sacs or force reality into artificial boxes that makes no sense. As Gielen (2019) explains in her work on terrorism and political violence, theory-driven evaluation mainly examines the theory of change that contributes reasonable hypotheses but

fails to provide empirical testing. Process-oriented evaluation focuses on implementation of the intervention, whereas realist evaluation examines context-mechanism-outcome patterns (Gielen 2019, 1151). Gielen (2019, 1152) stresses that it is a question not of comparing methods but of avoiding falling into 'black boxes' and that non-linearity in these kinds of contexts should be realized and considered.

At the workshop, most of the learning took place in open-ended conversations on shared failures and successes based on a combination of people's experiences and newcomers' curiosity in a constant dialogue between participants and organizers. To claim that a modern learning environment must be open-ended and trustworthy is almost a platitude today, but it must nevertheless be said, because it can be confused with simplified social interaction learning that emphasizes that learning is all about an interactional space and relations between people and learning by doing, as taught for decades (Dewey 1938, Freire 1970). Moreover, according to Freire (1970) learning is mainly about the disruption of oppressive norms, which is reverberated in Bourdieu's 'symbolic violence' and cultural capital (Bourdieu 1977). Learning in this environment becomes all about power and identity, leading to identity politics and fragmentation of society. In this kind of learning environment, disagreement is rare, and discussions about uncomfortable topics are not welcome.

Instead, we addressed facts from practice-based cases and scenarios and connected them to relevant concepts in a factual learning environment as Christodoulou (2014) has promoted. We could have done more on that. We realized that there is much to explore when it comes to teaching and learning how to evaluate in the context of FCV. Although the current level of evaluation knowledge in the field of FCV is scattered and insufficiently tested, as mentioned above, constructivist learning methods, which are now re-evaluated, have dominated approaches to pedagogy for a long time, because the constructivist learning models stress social learning and power at the expense of factual learning. On the other hand, Freire's (1970) emphasis on oppression and deconstruction of norms is relevant for teaching about evaluation in FCV, because of the intricate sociopolitical and cultural-religious web of relationships between victims and perpetrators that often involve economic dependency. This web is 'context', which is a key concept but is so deep that it requires not only more time and trust than the present evaluation designs in FCV provide, but also another pedagogic approach if any transformational effects are to be achieved.

Learning methods for FCV evaluation must be rethought and tested. We feel that we must leave the 'nice learning environment', with its clear

rules, patterns, repetitions and immediate feedback, and move into a 'sneaky learning environment', with its unpredictable rules and open-endedness (Kahneman and Klein 2009; see also Jelmini 2020). This sounds harsh, but it would prepare evaluators for real-world encounters in which, for example, trust must be earned and critically examined if we aim at transformational change with strong participatory dimensions. International and local evaluators are triggers for transformation in these contexts, assuming that the affected people find it worthwhile and reasonable to participate and learn. We need to jointly develop a learning environment for evaluators in FCV. For example, we should neither give way to slack participatory methods such as a romanticizing of local knowledge nor use participatory methods to manipulate stakeholders. An honest, realistic learning paradigm is required if we are to leave no one behind and be inclusive, as Van den Berg, Magro and Salinas Mulder (2019a) argue. We must face any ideological conformity in our teaching methods; put the facts on the table and rely on a pedagogy that aims at rational inquiries, inspirational solutions and no safe spaces.

Furthermore, to improve learning methods, we must systematize evaluation in FCV and conduct a scientific review of presently used evaluation methods and theories, similar to Gielen's (2019) previously mentioned study about countering violent extremism. The authors of this chapter have initiated such a review in collaboration with the IDEAS thematic group on FCV, including involuntary resettlement.

Further Thoughts on Evaluation in Wicked Environments

The authors of this chapter claim that evaluation in FCV has a higher complexity level than evaluation under normal circumstances. This is not to be interpreted that we undervalue the complexity of evaluation in non-violent environments, but FCV is fluid, violent, dangerous and unstable, and evaluation in FCV takes place in politically charged environments.

Transformational change in development entails the following: 'peace requires a complete transformation from situations of conflict and violence to sustainable peaceful relations between warring factions and societies' (Van den Berg, Magro and Salinas Mulder 2019a, 6). This has been discussed in the peace and conflict research literature since at least the classic works of Galtung (1969) – a pioneer in the field. A signed peace agreement is just the beginning of a long process of reconciliatory measurements and projects before a possible sustainable peace might be achieved that can

be called transformational. Likewise, there are several 'realities' in a conflict that are interconnected in time and space, and the adversaries are heterogeneous (Kriesberg 2015, 7–11). Therefore, an understanding of *when* and under *what* circumstances a complete transformation takes place and *who* is involved. These elements are essential for FCV evaluation and evaluation research. The problem is that evaluation designs in FCV contexts, if tested in a given environment, may not always work in another.

Apart from the evaluation designs suggested previously, one possible evaluation design is Feinstein's suggestion of a dynamic evaluation with learning loops that he argues is necessary for transformation. The learning loop consists of four causally interrelated elements: dynamic evaluations of interventions → policy dialogue → policy change → transformational change → and back to the beginning. This loop assumes that transformational change will be triggered and accountability supported (Feinstein 2019, 26). First, it is not clear where in the societal structures this transformational change occurs. Second, Feinstein assumes that policy matters and that a trickle-down effect occurs during implementation. This is a problem. In involuntary resettlement, policy safeguards (e.g. the World Bank's) have been applied since the 1980s, but research evaluation has shown that there is a worrisome gap between policy and implementation. The reasons are multidimensional.

The authors of this chapter are concerned that, in this cloudy nexus of relationships in FCV, there is a real risk of harm because of the fluidity and blurriness of context and stakeholders. Coherency, in the sense of systematic connection between parts, is a main difficulty in FCV evaluations. Maybe the only regularities are ad hoc events, fluid contexts and solutions, and if that is the case, evaluation must learn how to address those kinds of processes in a systematic way. This is not only a methodological and theoretical question, but also a question of legitimacy. It is about the use of resources for implementation of the SDGs in a meaningful way, as formulated by the OECD (2018, 13–14).

The COVID-19 Global Pandemic and FCV Evaluation

Hassnain et al. (forthcoming) argue that the COVID-19 pandemic is a case of FCV evaluation because the economic downturn of lower- and middle-income countries is likely to lead to an increase in violence and political instability. The Institute for Economics and Peace, in its report on

COVID-19 and Peace (2020), stated that most indicators in the Global Peace Index are expected to deteriorate. The one area that may improve is military expenditures, as countries redirect resources to propping up their economies.

The theme of the IDEAS Prague Conference was sharing experiences between the Global South and the Global North, which is in line with the idea of collaboration in FCV evaluation to enhance capacity, and this is why we bring cases from the Global North and the Global South.

In the Global South, the coronavirus has in many instances united people in raising their voices and coming out onto the streets in protest of how the states are handling such cases. For examples, millions of Brazilian protesters in São Paulo and Rio de Janeiro demonstrated against the government's handling of the coronavirus pandemic. They protested and called for the president to step down.

Prison breaks were reported in Venezuela, Brazil and Italy, with inmates reacting violently to new restrictions associated with COVID-19. Drug trafficking and other types of crime have seen a temporary reduction as a result of social isolation around the world, but reports of domestic violence, suicide and mental illness have increased (Institute for Economics and Peace 2020).

In the Global North, the scenario has been different. For example, in Sweden, the authorities did not lock down the country but decided to rely on *individual responsibility*. People were asked to work at home, avoid public transportation, socially distance, wash hands and stay at home if they felt sick. No punishments or reprisals were given. The strategy was intended to flatten the curve of infection so that the medical system would have time to adapt. It might also have encompassed the idea of achieving herd immunity for the population. This strategy might have resulted in a higher percentage of deaths than in neighbouring Nordic countries². Future evaluation research will show whether this was the case. As in the rest of the world, the distribution of people with COVID-19 is uneven, and state epidemiologist Anders Tegnell stated that 'This is an illness with very strong socio-economic links' (Omni 2020). The detailed reasons for this will be analysed in future

² Folkhälsomyndigheten (Public Health Agency of Sweden) provides the country with daily statistical updates and information on the distribution of documented COVID-19 deaths and the number of infected people distributed over the country according to region, urban versus rural districts, and male versus female. The statistics are public and transparent. Calculation mistakes are acknowledged and corrected in public, as is the fact that COVID-19 test kits imported from China have been inadequate. Affected people are informed and asked to test again (Folkhälsomyndigheten 2020).

research. Here we will only conclude that Sweden is not an FCV evaluation context because it has functioning multi-actor policy processes in place that intercepted the problems, one after another, and a large, educated population that more or less followed the rules. These multi-actor policy frames have been built up over a long time and are resilient.

On the other hand, an FCV context lacks stability, although stability per se also needs further analysis, as discussed in the peace and conflict literature. The lack of stability (instability?) is combined with a small, educated elite and a large uneducated population who are used to corruption, lack of transparency and government incompetence, or at least a perception that this is the case. Without stability (sustainable socioeconomic-political and cultural structures) and trust, a country faces immense difficulties in the fight against a pandemic.

An effective FCV evaluation must explicitly address these issues in an unsentimental, non-ideological, rational way and not uncritically lean on well-meant general concepts repeated over and over in research and evaluation. For example, let us bring up the concept of 'local', which is usually approached as a level but maybe is better seen as 'a standpoint based in a particular locality, but not bounded by it', as by Shaw and Waldorf (2010, 6) suggested. This is context, and we must cut through its complexity. Maybe a way forward is, as Feinstein (2019, 21–25) argues, a dynamic evaluation with its key aspects of relevancy, multiple methods, scaling up, quasi-real-time evaluation and political sensitivity. The dynamic approach is valid because of the fluidity of an FCV context, and further experimentation to assess its compatibility with the suggested methodological elements is needed. Alternatively, what might be needed is a system thinking evaluation approach (Magro and Van den Berg 2019) that can incorporate and make use of fine-meshed qualitative data in a grid of quantitative data. There is no use in continuing with 'meaning-making' studies in an FCV context, if they continue to fail to operationalize their important messages, especially in an FCV context. It could even be dangerous. In the following section, we will discuss qualitative data that are systematized with the help of the realistic evaluation model using key concepts.

Violent Extremism and Evaluation Methods

Exploring further the authors' claim that evaluation in FCV is special because of the context, we will examine closely a case of evaluating programmes for countering violent extremism in Europe. This case study could help identify

possible further criteria for FCV evaluations or, should we say, evaluation in FCV contexts.

Gielen (2020), in her doctoral thesis 'Cutting Through Complexity', has investigated how to evaluate countering violent extremism in Europe. In two earlier articles (Gielen 2018; 2019), she developed and confirmed her argument. Here we will focus on the article on how to evaluate female jihadist exit programmes in the Netherlands (Gielen 2018).

When the Islamic State in Syria (ISIS) proclaimed its caliphate in June 2014, it led to an increase in female recruitment from European countries – young women who travelled to Syria and Iraq to marry ISIS fighters. They became known as 'jihadi brides' to become mothers to 'cubs of the caliphate'. In Europe, they were seen as victims, and Gielen (2018) argues that this 'victimization' prevented effective policies against female violent extremism. Recent empirical studies have shown that female jihadists are far from victims, and this requires a more nuanced view of the returning female jihadists in order to develop effective counter-measure programmes of deradicalization and reprogramming. These programmes must be tailor-made to be successful, and this is a challenge to prevailing evaluation methodologies. Gielen suggests that realistic evaluation is an option to find out what works for whom and when. This is also in line with our experiences with FCV, as discussed earlier.

Exit programmes consists of several elements, such as deradicalization (changing extremist beliefs), disengagement (dissuading from violent extremist action), reintegration and rehabilitation. All of these elements, if implemented, can lead to transformational change for the individual and for society.

Their deconstruction, which starts with their binary opposition, makes the elements mentioned above operational; for example, deradicalization starts with an investigation of radicalization, in which push and pull factors were identified, such as roles and gender (Gielen 2018). Female jihadists' gender roles were often underestimated, and they became an unrecognized and dangerous threat. Any exit programme must consider this.

The problem with the existing exit programmes was that, as Demant and colleagues say, 'exit programmes for jihadists focus too much on normative factors, concentrating on theological and ideological issues, and as a consequence overlook affective factors such as the family and peer network' (quoted in Gielen 2018, 460). This is also in agreement with our approach to FCV – to leave the frontstage context and reach the backstage context to achieve a kind of holistic view while still keeping the ethnographic details in mind.

Furthermore, Gielen (2018) argues that the local level was crucial for the understanding and success of the exit programme. Here, 'local level' means municipalities' multi-agency resources and management, such as legal, administrative and 'soft' resources (psychological counselling, family support, practical support with job and housing) and help with breaking with extremist networks, because grooming is one of the biggest problems. When a female jihadist returns, she is contacted within a short time (48 hours) to be convinced or threatened to rejoin the network.

Hence, 'local' comes back into the evaluation model as a key concept and engages with multi-actor agencies in a processual way. In Gielen's final model of a successive exit programme for female jihadists, several well-established, key elements return. She stresses a dense contextual approach. The municipality and its institutional setting constitute the frame of this context. Within this frame, gender perspectives and demography are taken into account and linked to mentoring programmes to promote trust and establish a long-term, stable environment. The stability depends on the quality of the relations between the agencies in charge of the case. Furthermore, concrete legal measures are applied to prevent the women from rejoining the network. Such legal measures can include confiscating passports, prohibiting contacts and banning social media (Gielen 2018). Pawson and Tilley (1997, 40) argue that, to establish an effective exit programme, multi-method data collection is necessary but without falling into 'the experimentalist trap to compare "inputs" and "outputs" in the sense that some programmes do work and others do not. Rather, realist evaluation should concern itself with the "make-up" of the interventions and respondents to address the question why some programmes work better for some than for others.'

Whether the above reasoning would effect transformational change is an open question, but it might be, as Feinstein (2019) argues, that to be transformative, evaluation must change focus from projects and programmes to strategies and policies. In addition, gender researchers argue that systemic gender knowledge is a presumption of transformational change, because 'scholars agree that gender inequality is systemic and that participants in gender equality interventions need knowledge on gender inequality processes' (Lansu, Bleijenberg and Benschop 2019, 1589). From an FCV evaluation perspective, we do not know what the triggers are for transformational change in any systematic way, and we are painfully aware of the gap between policy and implementation. Even more difficult, we have not pinpointed an exact moment in time when we can say whether an intervention has caused more harm than good.

Conclusion

Evaluation, like development aid, can unintentionally exacerbate tensions in ways that harm conflict-affected populations if care is not taken to develop and integrate layers of sensitivity into the design and approaches and throughout the evaluation cycle.

The contexts of FCV pose particular challenges for evaluation, such as difficulties in accessing the affected population; limited availability of good-quality data; lack of appropriate tools and resources; challenges in navigating conflicted stakeholder and informant relationships; high levels of unintended effects; signs of corruption and human rights violations that are difficult to validate and report on and political volatility that makes it difficult to identify key stakeholders of the intervention to be evaluated. These difficulties, if not properly addressed or mitigated, can call the validity of the evaluation into question. Mitigating them can be an important challenge and sometimes requires redefining the direction, purpose or scope of the evaluation, which requires that not only the evaluation team, but also the donor or funding agency be flexible.

Although evaluation data collection methods continue to evolve, there is a need for a more comprehensive repository that assembles guidance from different sources to facilitate learning about different approaches to conducting evaluation in the context of FCV in climate- and gender-sensitive ways. These are politically volatile environments, and it can be difficult to identify the key stakeholders that must be on board for an authentic evaluation process because the narration and understanding of conflict is different for each party or institution. It can be challenging to find the appropriate direction, purpose and scope for the evaluation, and sometimes this has to be redefined when in the field.

In FCV contexts, the importance of understanding the cultural, socio-economic and political context assumes a higher importance given the inherent complexities of such contexts. This includes not only facts about the conflict, but also the culture, economy, sociopolitical structures and safety – understanding in its true sense. This may be impossible with the present routines of evaluation missions. Nevertheless, this is what we have to work with. The authors emphasize the importance of conducting an in-depth desk study, including a detailed evaluability assessment to consider whether the evaluation is feasible and appropriate in the given context. This is also helpful in many ways to clarify data gaps and any other operational evaluation matters. In addition, the MDG progress reports indicate that data collection in FCV environments has been poor or that data

on the indicators are unavailable on numerous occasions. Some of the key reasons behind this is the absence of evaluation systems, tools and opportunities. The Global Evaluation Initiative verifies this need to harmonize tools and resources for effective evidence generation at all levels, indicating that only one-third of the countries that committed to achieving the SDGs have sufficient monitoring and evaluation capacities to measure any changes on the ground.

The bottom line is that evaluating in FCV contexts always starts with investing time and resources in analysing the context, including the key drivers of conflict and how these affect the rest of the society, including the interventions designed in such contexts, and vice versa. It will not prevent evaluations from missing or poor-quality data or, more importantly, from making expensive mistakes. We must learn to be imperfect, although we must strive towards perfectness and, foremost, help evaluators ensure learning and promote accountability at all levels. Just beware of the fact that, in normal circumstances, failures bring learning, whereas failure in proper planning and executing an evaluation may mean serious life-threatening consequences.

Let us conclude with some final thoughts on these intricate issues. The impression is that evaluation for transformative change envisions an Eden world of truth, equality, justice and peace, echoing Bob Thiele's famous 1967 song, magnificently interpreted by Louis Armstrong, 'What a Wonderful World'. The problem is how to reach this wonderful world, and the path is stacked with obstacles. As we see it, transformation is about memory but also about forgetting, and maybe (but only maybe) this is particularly important in fragile and violent societies. Transformation is about how and what to remember and how and what to forget. To reconcile, or at least to go on, is to be able to hear the sound of things falling, in the words of Juan Gabriel Vásquez (2011). It is also about not repeating 'This way for the gas, ladies and gentlemen', following the book title of holocaust survivor Tadeusz Borowski (1967).

Feinstein (2019) is referring to transformation in his review of the famous novel *Il Gattopardo* (*The Leopard*), by Giuseppe Tomasi di Lampedusa, published posthumously in 1960. Feinstein takes Lampedusa's famous quotation – 'Everything must change for everything to remain the same' – literally (to change to be the same) and regards it as a micro-event that does not really influence the larger societal change. We think this is a misinterpretation, because on Don Fabrizio's deathbed, the quotation becomes 'if you don't change, time will change you' (di Lampedusa 1960), or in the words of the Museum of di Lampedusa, 'It is a novel where the temporal

limits of human nature are always present, melancholic, touching and wise' (Butera 28 Apartments 2020). As we see it, transformation is about temporal and spatial orders and the human consciousness.

Evaluating for transformational change in unstable and fragile circumstances is to cut through complexity. We believe that evaluation and evaluation research must be a player in this matrix, but we need to recognize, grasp and learn how to use fewer tangible elements of change to make it happen, as the authors of this chapter have discussed in an earlier text (Aronsson and Hassnain 2019) on value-based evaluation for transformative change.

References

- Aronsson, I-L., and H. Hassnain. 2019. Value-based evaluations for transformative change. In: *Evaluation for transformational change: Opportunities and challenges for the Sustainable Development Goals*, (eds.) R.D. van den Berg, C. Magro and S. Salinas Mulder, 89–104. Exeter, UK: IDEAS.
- Aronsson, I-L., and H. Hassnain. Forthcoming a. Ethics in evaluation in resettlement related interventions. In: *Evaluation in an ethically challenged world*, (eds.) R.D. van den Berg, P. Hawkins and N. Stame. Exeter, UK: IDEAS.
- Aronsson, I-L., and H. Hassnain. Forthcoming b. Evaluation and ethics in contexts of fragility, conflict and violence. *Evaluation in an ethically challenged world*, (eds.) R.D. van den Berg, P. Hawkins and N. Stame. IDEAS
- Bamberger, M. 2007. *RealWorld Evaluation: Working under budget, time, data and political constraints*. Oregon Program Evaluators Network Brown Bag Series, 27 September.
- Bennett, J., S. Pantuliano, W. Fenton et al. 2010. *Aiding the peace: A multi-donor evaluation of support to conflict prevention and peacebuilding activities in southern Sudan 2005–2010*. Hove, UK: Itad Ltd.
- Borowski, Tadeusz. 1967. *This way for the gas, ladies and gentlemen*. London: Penguin Books.
- Bourdieu, Pierre. 1977 [1972]. *Outline of a Theory of Practice*. Trans. Richard Nice. Cambridge, UK: Cambridge University Press.
- Butera 28 Apartments. 2020. The Leopard – Giuseppe Tomasi di Lampedusa. Webpage. <https://tinyurl.com/4sbw3exu>
- Chigas, Diana et al. 2006. *Has peacebuilding made a difference in Kosovo? A study of the effectiveness of peacebuilding in preventing violence: Lessons learned from the March 2004 riots in Kosovo*. Case Study, Cambridge, MA: CDA Collaborative Learning Projects and Care International. <https://tinyurl.com/4yawunck>
- Christodoulou, D. 2014. *Seven myths about education*. London, UK: Routledge.

- Dewey, J. 1938. *Experience and education*. West Lafayette, IN: Kappa Delta Pi.
- Di Lampedusa, G.T. 1960. *Leoparden (Il Gattopardo)*. Swedish translation. Bonnier: Stockholm.
- Easterly, W. 2015. The SDGs should stand for senseless, dreamy, garbled. *Foreign Policy* 28 September. <https://tinyurl.com/y21mtp2x>
- Feinstein, O. 2019. Dynamic evaluation for transformational change. In: *Evaluation for transformational change: Opportunities and challenges for the Sustainable Development Goals*, (eds.) R. D. van den Berg, C. Magro and S. Salinas Mulder. Exeter, UK: IDEAS.
- Folkhälsomyndigheten (Public Health Agency of Sweden). 2020. COVID-19 statistics. <https://tinyurl.com/49d5a4jr>
- Freire, P. 1970. *Pedagogy of the oppressed*. Penguin Classic.
- Galtung, J. 1969. Conflict as a way of life. *Essays in Peace Research* 3: 484–507.
- Gielen, A.-J. 2018. Exit programmes for female jihadists: A proposal for conducting realistic evaluation of the Dutch approach. *International Sociology* 33 (4): 454–472.
- Gielen, A.-J. 2019. Countering violent extremism: A realist review for assessing what works, for whom, in what circumstances, and how? *Terrorism and Political Violence* 31 (6): 1149–1167.
- Gielen, A.-J. 2020. Cutting through complexity. Evaluating countering violent extremism (CVE). Doctoral thesis. University of Amsterdam.
- GEI (Global Evaluation Initiative). 2020. Global Evaluation Initiative. <https://www.globalevaluationinitiative.org/>
- Hassnain, H. 2017. Measuring peace in our lifetime: Evaluating SDGs. *Alliance Magazine*. <https://tinyurl.com/yfmbh5>
- Hassnain, H., L. Kelly and S. Somma, eds. 2021. *Evaluation in contexts of fragility, conflict and violence: Guidance from global evaluation practitioners*. Exeter, UK: IDEAS.
- Hassnain, H., and M. Lorenzoni. 2020. Embracing the pandemic. Institute for Economics and Peace. 2017. SDG 16 progress report. Sydney.
- Institute for Economics and Peace. 2020. COVID-19 and Peace. Sydney.
- Jelmini, M. 2020. Tidig specialisering inte recept på framgång – snarare tvärtom. *Svenska Dagbladet* 13 August 2020, 26–27.
- Kahneman, D., and G. Klein. 2009. Conditions for intuitive expertise. A failure to disagree. *American Psychologist* 64 (6): 515–526. <https://doi.org/10.1037/a0016755>
- Kelly, L., D. Nogueira-Budny and J. Chelsky. 2020. Conflicting results: Measuring outcomes in situations of conflict. World Bank blogs. <https://tinyurl.com/s2fennk>
- Kriesberg, L. 2015. *Realizing peace. A constructive conflict approach*. Oxford: Oxford University Press.
- Lansu, M., I. Bleijenberg and Y. Benschop 2019. Seeing the system: Systemic gender knowledge to support transformational change towards gender equality in science. *Gender Work Organization* 26: 1589–1605.

- Magro, C., and R.D. van den Berg. 2019. Systems evaluations for transformational change. In: *Evaluation for transformational change: Opportunities and challenges for the Sustainable Development Goals*, (eds.) R.D. van den Berg, C. Magro and S. Salinas Mulder, 131–152. Exeter, UK: IDEAS.
- OECD (Organisation for Economic Co-operation and Development). 2018. *States of fragility 2018*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264302075-en>
- Omni. 2020. Undersökning: Snabbast spridning i utsatta områden. <https://tinyurl.com/3z648np9>
- Oxfam. 2020. Oxfam report. <https://tinyurl.com/56vmc8th>
- Pawson R., and N. Tilley. 1997. *Realistic evaluation*. London: Sage. <https://tinyurl.com/rcazb76v>
- Shaw, R., and L. Waldorf, with P. Hazan (eds.) 2010. *Localizing transitional justice. Interventions and priorities after mass violence*. Stanford, CA: Stanford University Press.
- Van den Berg, R. D., C. Magro and S. Salinas Mulder. 2019a. Introduction: Bringing challenges for evaluation for transformational change from Guanajuato to Prague. In: *Evaluation for transformational change: Opportunities and challenges for the Sustainable Development Goals*, (eds.) R.D. van den Berg, C. Magro and S. Salinas Mulder. Exeter, UK: IDEAS.
- Van den Berg, R. D., C. Magro and S. Salinas Mulder (eds.). 2019b. *Evaluation for transformational change: Opportunities and challenges for the Sustainable Development Goals*. Exeter, UK: IDEAS.
- Vásquez, J.G. 2012. *The sound of things falling*. New York: Riverhead Books.
- World Bank. 2017. Fragility, conflict and violence. <https://tinyurl.com/3tf7a3hd>

The COVID-19 pandemic has demonstrated the enormous challenges humanity is facing. It has been facilitated by other crises as climate change, biodiversity loss, economic exploitation, and increased inequity and inequality. The UN Agenda 2030 and the Paris Agreement on climate change call for transformational change of our societies, our economies and our interaction with the environment. Evaluation is tasked to bring rigorous evidence to support transformation at all levels, from local to global. This book explores how the future of the evaluation profession can take shape in 18 chapters from authors from all over the world, from North and South, East and West, and from Indigenous and Decolonized voices to integrative perspectives for a truly sustainable future. It builds on what was discussed at the IDEAS Global Assembly in October 2019 in Prague and follows through by opening trajectories towards supporting transformation aimed at solving the global crises of our times.

By combining practical experiences with perspectives drawn from new initiatives, this book offers invaluable insights into how evaluation can be transformed to support transformational change on the global stage.

Indran A. Naidoo, Director of the Office of Independent Evaluation of IFAD

Across continents, educational systems, and historical complexities, this book builds up the language we all should speak about our field. A mandatory read for all young evaluators.

Weronika Felcis, Board member of EES and Secretary of IOCE

After reading these chapters you will have a sharper look at what is relevant when managing or doing an evaluation, and you will notice that 'business as usual' will no longer be an option.

Janett Salvador, Co-founder of ACEVAL, Former Treasurer of ReLAC

This book offers original, visionary discourse and critical perspectives on the challenges evaluation is facing in the post COVID-19 pandemic era.

Doha Abdelhamid, Member of the Egyptian Academy of Scientific Research and Technology

Published by: IDEAS, 2021
ISBN (paper): 978-1-9168982-0-2
ISBN (electronic): 978-1-9999329-9-2

