



The effectiveness of peacebuilding evaluation: Facing the attribution gap with research methods

With a case study on the attitudes, perspectives, and experiences of six organizations that execute and/or evaluate peacebuilding interventions

Maartje Pott
S1027267

MSc Political Science
Specialization Conflict, Power and Politics

Nijmegen School of Management
Radboud University, Nijmegen, The Netherlands

Title: The effectiveness of peacebuilding evaluation: Facing the attribution gap with research methods

Subtitle: *With a case study on the attitudes, perspectives, and experiences of six organizations that execute and/or evaluate peacebuilding interventions*

Name: Maartje Pott

Student number: S1027267

Education program: Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master in Political Science (MSc)

Specialization: Conflict, Power and Politics

Supervisor and first reader: Dr. ir. M. van Leeuwen

Second reader: Dr. W. Plowright

Faculty: Nijmegen School of Management

University: Radboud University, Nijmegen, The Netherlands

Date of submission: June 24th, 2021

Words: 17.021 (excluding references and appendices)

'I think it's always difficult to say, OK, this happened thanks to us and it's a bit presumptuous. Right?'

Participant 1, Interview 1

PREFACE

After completing the BSc International Development Studies at Wageningen University, I wanted to learn more about (inter)national power relations and decided to dive into the Political Science field at Radboud University. After the pre-MSc Political Science, I decided to focus on the dynamics behind conflict and peace and therefore chose the Conflict, Power and Politics track of the MSc Political Science.

This thesis consists of a case study on effective peacebuilding evaluation and, besides my great interests, has been written to meet the graduation requirements for the Degree of Master in Political Science (MSc) (Conflict, Power and Politics) at Radboud University, Nijmegen, the Netherlands. This thesis was carried out from the end of January 2021 up until the end of June 2021.

I first would like to thank my kind supervisor dr. ir. M. van Leeuwen, who provided me with a lot of knowledge, ideas, suggestions, and guidance throughout the process from narrowing down the research topic up until and including the eventual thesis laying before you. Additionally, I would like to thank the organizations that were willing to share their experiences and perspectives with me.

I hope you will enjoy reading this thesis.

Yours sincerely,
Maartje Pott

Zwolle, June 21st, 2021

ABSTRACT

The peacebuilding evaluation field stumbles upon challenges. One of those challenges is the attribution gap. In practice, it turns out to be difficult to attribute an intervention to (a) societal change(s) or impact(s). To make peacebuilding evaluation more effective, it can be argued that this gap can be resolved by using qualitative research methods within peacebuilding evaluation, since this approach is better able to capture the various (changing) external variables within the context of the intervention than quantitative research methods according to theory. To test this hypothesis, the attitudes, perspectives, and experiences of six organizations that execute and/or evaluate peacebuilding interventions on effective peacebuilding evaluation, the attribution gap, and the use of quantitative and qualitative research methods within peacebuilding evaluation are analysed through a deductive descriptive case study. Through semi-structured interviews, this research finds: (1) there are various definitions of effective peacebuilding evaluation; (2) evaluation and resolving the attribution gap are not always a priority; and (3) not all of the organizations are able to describe what their evaluation methods are. Nonetheless, the research finds ambitions on wanting to capture impact. Quantitative methods are perceived as not comprehensive enough to fulfil this task in a peacebuilding setting, while qualitative methods are seen as a necessity to measure this. Follow-up research could shed more light on: (1) the lack of consistency in defining effective peacebuilding evaluation; (2) the lack of knowledge on what evaluative methods are used; and (3) why ambitions do not match practices; before diving into the attribution gap.

Contents

List of abbreviations and acronyms	7
List of tables	8
Introduction.....	9
Theoretical framework.....	13
Methodology.....	17
Chapter 1: What criteria do the contributing organizations use to assess effective peacebuilding evaluation?	22
Chapter 2: In what ways and to what extent do the contributing organizations experience the attribution gap?	26
Chapter 3: Do the contributing organizations prefer certain research methods within peacebuilding evaluation to capture the impact of their peacebuilding interventions?.....	29
Conclusion.....	32
Discussion	35
References	38
Appendix I: Semi-structured interview protocols	40
Appendix II: Coding scheme.....	41

List of abbreviations and acronyms

CR/PB	Conflict Resolution/Peacebuilding
IOB	Policy and Operations Evaluation Department
OECD	Organisation for Economic Co-operation and Development
M&E	Monitoring and Evaluation
QUANT	Quantitative
QUAL	Qualitative

List of tables

Table 1 Revised Evaluation Criteria Definitions and Principles for Use	14
Table 2 Key differences between qualitative and quantitative research	15
Table 3 Characteristics of the contributing organizations	20
Table 4 Members of the OECD.....	23
Table 5 Overview of attitudes, perspectives, and experiences of the contributing organizations.....	34

Since 1946, the number of conflicts has increased (Roser, 2016). In response to this trend, peacebuilding became ‘[...] by far the predominant form of multilateral peace operation [...],’ (Barma, 2017, p. 12). With around 33 armed conflicts in 2020 (IISS, 2020), peacebuilding continues to be part of the agenda of, amongst others, governments, intergovernmental bodies, and non-governmental organizations (Çuhadar-Gürkaynak, Dayton & Paffenholz, 2009).

While peacebuilding plays an important role within the response to violent conflict, there are doubts about its effectiveness; peacebuilding sometimes seems to cause more harm than good (Autesserre, 2017; OECD, 2012a). To capture the effectiveness of a peacebuilding intervention, the peacebuilding evaluation field provides tools to measure the actual effect(s) of the implemented intervention. Albeit that is what it supposed to do.

Peacebuilding evaluation

When trying to execute effective peacebuilding evaluations, the evaluation field stumbles upon many challenges which limit the ability to execute the evaluations properly. According to Çuhadar-Gürkaynak et al. (2009) the greatest challenges that evaluators face when trying to evaluate peacebuilding interventions relate to: *theories of change, the attribution gap, the gap between the evaluation preferences of donors and local stakeholders, and the difficulties surrounding data collection in conflict-affected areas*. It is important to note that these challenges are identified as the main obstacles but are by no means exhaustive. Additionally, these challenges relate to each other and can overlap. (Çuhadar-Gürkaynak et al., 2009)

The first challenge is related to the theories of change, which are the starting point of peacebuilding evaluation. It is ‘[...] a set of beliefs about how change happens and, as such, it explains why and how certain actions will produce the desired changes in a given context, at a given time,’ (OECD, 2012a, p. 29). However, *theories of change tend to be not thoroughly articulated at the start of peacebuilding practices*. This is due to the difficulty of connecting the dynamics of a conflict to proper peacebuilding interventions. This means that there is no explicitly articulated fit between the two. Eventually, the lack of a thoroughly articulated theory of change at the beginning of a peacebuilding practice means that it is not clear ‘[...] what they [red. conflict resolution/peacebuilding managers] are doing, why they are doing it, and how they can determine whether or not their objectives have been achieved once the activity has been concluded,’ (Çuhadar-Gürkaynak et al., 2009, p. 293). Without a proper theory of change, it is therefore not clear what to evaluate specifically. The participants of the Peacebuilding Evaluation Summit also experience difficulties with the theories of change (Blum & Kawano-Chiu, 2012). According to the participants these theories are not effective when they are only based upon project evaluations. Blum and Kawano-Chiu (2012, p. 10) show that these evaluations mostly do capture if a project was effective or not, but not if ‘[...] a broader peacebuilding strategy did not work [...]’. The participants argue that the effectiveness of theories of change would increase if ‘[...] basic information on social dynamics and social change,’ was included (Blum & Kawano-Chiu, 2012, p. 10). However, due to the lack of this basic information, theories of change are mostly based on project evaluations. This brings the whole challenge around theories of change full circle. (Blum & Kawano-Chiu, 2012; Çuhadar-Gürkaynak et al., 2009)

The second difficulty within the evaluation of peacebuilding relates to *the attribution of complex social impacts to a peace effort*. Çuhadar-Gürkaynak et al. (2009) call this the attribution gap. Measuring social phenomena is difficult in social science in general due to the inability to rule out external factors. However, measuring the effect of a peacebuilding intervention within conflict-affected areas makes it even more complex, due to the ‘[...] multitude of uncontrollable external variables [...]’ present in a conflict setting on micro and macro level (Çuhadar-Gürkaynak et al., 2009, p. 295). They explain: ‘For instance, economic hardships, shifting regional political alliances, leadership transitions, or even the activities of diaspora communities thousands of miles away can directly or indirectly impact the ability of the CR/PB initiative to meet its objectives,’ (Çuhadar-Gürkaynak et al., 2009, p. 295). In other words,

the economic, political, social and/or organizational influences and changes make it difficult to measure if an intervention has been effective. Scharbatke-Church (2011), OECD (2012b), and Anderson and Olson (2003) also point towards the attribution challenge within peacebuilding evaluation.

Evaluation preferences between donors and the local community in the conflict-affected areas are not always in line with each other. Çuhadar-Gürkaynak et al. (2009, p. 296) explain that this is due to the differences in goals of both parties within an evaluation. Donors mostly focus on '[...] efficiency, timeliness, sustainability, and coherence [...]' and '[...] predetermined and verifiable indicators of change [...]', while local citizens in the conflict-affected areas rather focus on the urgent needs of the people in these areas like '[...] ceasing hostility, providing safe havens for refugees, or delivering essential services' (Çuhadar-Gürkaynak et al., 2009, p. 296). Although there are evaluation approaches that try to include all stakeholders, and therefore try to include all these different voices when carrying out an evaluation, in practice it is experienced as difficult to carry out evaluations effectively when one wants to measure such different indicators. Therefore, the inclusion of local voices in peacebuilding evaluation is sometimes lacking (Çuhadar-Gürkaynak et al., 2009).

The last challenge many peacebuilding evaluators face is *the difficulties that come with data collection in conflict-affected areas*. It is sometimes impossible or too dangerous. The attendees of the Peacebuilding Evaluation Summit additionally highlight the risks involved with data collection. They for example argue that it can affect: the effectiveness of the peacebuilding program, the dynamics in the community and the values of the organization. (Blum & Kawano-Chiu, 2012; Çuhadar-Gürkaynak et al., 2009)

Narrowing the scope

Çuhadar-Gürkaynak et al. (2009) have tried to provide solutions to the challenges presented above. The solution they suggest to resolve the attribution gap seems the most puzzling. They propose that practitioners executing evaluations must '[...] limit claims about the impacts of particular CR/PB activities to those that can be validated through carefully designed evaluative methods,' (Çuhadar-Gürkaynak et al., 2009, p. 296). But how do you carefully design evaluation methods that capture the various 'uncontrollable external variables' (Çuhadar-Gürkaynak et al., 2009, p. 295)? What evaluative methods are useful to address the attribution gap?

To figure out how to capture these various variables, it could be argued that it is needed to understand the context of the peacebuilding intervention since this is where these variables are present. According to Denskus (2012), the monitoring and evaluation field of peacebuilding is however not able to capture the complexity of a conflict-affected area and the impact of the peacebuilding practices due to the dominant use of quantitative research methods. Quantitative research methods within peacebuilding evaluations do not '[...] engage participants or raise fundamental questions' that are needed for effective evaluation (Denskus, 2012, p. 151). The use of mostly quantitative methods within peacebuilding evaluations is confirmed by the Development Assistance Committee (2007). The committee argues that the indicators that are evaluated within peacebuilding '[...] are rarely accompanied by qualitative assessments of the operations under evaluation or their context,' (Development Assistance Committee, 2007, p. 27). The Committee furthermore states: '[...] there is an overemphasis on quantifiable indicators even though conflict prevention and peacebuilding often involves non-quantifiable outcomes and impacts,' (Development Assistance Committee, 2007, p. 28). Denskus (2012, p. 148) proposes:

'[...] rather than spending more time and effort on quantitative evaluation methodologies an increased interest in aid relationships (e.g. Eyben, 2006; Mosse and Lewis, 2005), qualitative methods and ethnographic research will strengthen monitoring and evaluation of peacebuilding work or aid work in 'difficult' environments and ultimately lead to increased learning possibilities for everyone involved.'

He states that qualitative research methods are better able to capture, amongst others, political, economic, and social frameworks within the society where the peacebuilding interventions are executed, because it sheds light on unperceived perspectives from people in the conflict-affected area.

To contribute to the debate on what ‘carefully designed evaluative methods’ should consist of to face the attribution gap and make peacebuilding evaluation more effective, more knowledge needs to be gathered on methods (Çuhadar-Gürkaynak et al., 2009, p. 296). Therefore, this thesis narrows its scope to gathering attitudes, perspectives, and experiences of organizations that execute and/or evaluate peacebuilding interventions on effective peacebuilding evaluation, with a focus on the attribution gap and the use of quantitative and qualitative research methods within their peacebuilding evaluations.

Relevance

Getting to know more about these attitudes, perspectives and experiences regarding effective peacebuilding evaluation, the attribution gap, and the use of quantitative and/or qualitative research methods might help to provide insight into how qualitative research methods can help to resolve the attribution gap. This research furthers not only this academic debate, it also adds to the academic debate on what effective peacebuilding evaluation means in the first place, by investigating the different perceptions upon effective peacebuilding evaluation. Additionally, by getting to know more about the perspectives regarding peacebuilding evaluation, it continues the debate on the value and necessity of peacebuilding evaluation. Moreover, providing insight into the different perceptions upon the attribution gap furthers the debate on the importance of resolving this gap.

Besides the academic relevance of this topic, it also has societal value. Getting to know more about the organizations’ views regarding effective peacebuilding evaluation shows where peacebuilding evaluation currently stands in practice. Moreover, collecting and analysing these different views provides organizations an opportunity to learn from other organizations, which could better future evaluations. In the long run, this increase in evaluation knowledge could eventually lead to better future peacebuilding interventions. More specifically, an increase in knowledge around methodological approaches, leads to an increase in ‘[...] learning and accountability promises’ of peacebuilding evaluation according to Scharbatke-Church (2011, p. 479). This research also furthers knowledge on how evaluations can become more cost-effective. The participants of the Peacebuilding Evaluation Summit note that ‘[...] sufficient funding is required to make methodologies more rigorous [...],’ (Blum & Kawano-Chiu, 2012, p. 8). Çuhadar-Gürkaynak et al. (2009, p. 288) also argue: ‘what methods will be used for an evaluation depends [...] [red. on] the size of the available budget.’ By expanding knowledge on effective peacebuilding evaluation methods, peacebuilding evaluation can be made more cost-effective.

Problem statement, purpose, and research questions

All in all, peacebuilding evaluations could become more effective if the attribution gap would be resolved. The solution seems to revolve around ‘carefully designed evaluative methods’ (Çuhadar-Gürkaynak et al., 2009, p. 296), which are argued to consist of qualitative research methods since these seem to be more comprehensive in measuring the complex context, impacts and outcomes that need to be understood to be able to make attribution claims (Denskus, 2012; Development Assistance Committee, 2007).

To contribute to more effective peacebuilding evaluation, the goal of this thesis is to capture the attitudes, perspectives, and experiences of six organizations that carry out and/or evaluate peacebuilding interventions on effective peacebuilding evaluation, with a focus on how they try to overcome the attribution gap through quantitative and/or qualitative research methods within their evaluations. This thesis tries to determine what ‘carefully designed evaluative methods’ mean to these six organizations, by analysing their preferences when they measure the effect of their interventions. Three out of six are international non-governmental organizations, of which one can be seen as a peacebuilding organization. The other two mostly focus on humanitarian and development interventions of which peacebuilding is a component. In addition, two of these six organizations are non-governmental organizations that are

commonly referred to as local partners. The remaining organization carries out evaluations mainly on behalf of the government.

These organizations remain anonymous within this thesis. This thesis does not want to single out challenges of particular organizations, since the challenges they face might be representative for a wider group of organizations that execute and/or evaluate peacebuilding interventions. Additionally, there are also power relations at play between the contributing organizations, like donor relationships or consortia. This thesis does not want to contribute to possible hardships between the organizations that could make donor relations or consortia vulnerable if identities or organization names would be known. Moreover, the interventions of the organizations are mostly executed in conflict-affected areas, which are highly political sensitive. Therefore, certain information about (an) organization(s) and/or (an) intervention(s) are anonymized.

The central research question of this thesis is the following:

How do organizations that execute and/or evaluate peacebuilding interventions perceive quantitative and/or qualitative research methods within peacebuilding evaluations as more satisfactory in addressing the attribution gap?

The sub-questions related to this question are:

1. What criteria do the contributing organizations use to assess effective peacebuilding evaluation?
2. In what ways and to what extent do the contributing organizations experience the attribution gap?
3. Do the contributing organizations prefer certain research methods within peacebuilding evaluation to capture the impact of their peacebuilding interventions?
 - 3.1 How are these research methods expected to help resolve the attribution gap?
 - 3.2 To what extent do these research methods in practice fulfil the criteria of the contributing organizations on effective peacebuilding evaluation?

Methodology

To answer these research questions, this thesis uses a deductive qualitative methodological approach. This means that the theoretical debates presented in this chapter and the following chapter are the starting point of the research that are tested by this thesis. The method used to collect data are semi-structured interviews through a descriptive case study on six organizations that carry out and/or evaluate peacebuilding interventions. To analyse the collected data a discourse analysis is used, which is both deductive and inductive, but mostly leans towards a deductive discourse analysis since most codes are built on the theoretical debates.

Structure

This thesis first provides the theoretical basis of this descriptive case study. This theoretical framework goes into the debates surrounding peacebuilding, peacebuilding evaluation, and preferences for particular (qualitative) research methods (in the evaluation field). After the methodological chapter, which explains the approach of this descriptive case study into depth, the results of the data collection are presented. Thereafter, in the concluding chapter, this thesis portrays its central findings, answers the research question and elaborates on how this case study contributes to the academic and societal debates. This is followed by the discussion, which sheds light on the limitations and the findings that fall outside the scope of this thesis. This chapter ends with recommendations for follow-up research and in what way peacebuilding evaluation could become more effective.

Theoretical framework

As an addition to the debates presented in the Introduction, this chapter presents the concepts and current debates regarding peacebuilding and peacebuilding evaluation. After the focus on these two topics, this chapter dives into the debate regarding effective peacebuilding evaluation research methods and the attribution gap.

Peacebuilding

Peacebuilding is a broad concept that changed over time. Autesserre (2017, p. 115) defines it as: ‘actions aimed at creating, strengthening, and solidifying peace (Boutros-Ghali 1992)’. Peacebuilding depends on the notion of peace underlying these actions. Peace can be defined as: negative peace or positive peace (Autesserre, 2017). Negative peace relates to ‘the absence of war and violence’ (Autesserre, 2017, p. 115). Peacebuilding interventions based on negative peace are e.g., ‘militia demobilisation programmes’ (Scharbatke-Church, 2011, p. 463). Negative peace ‘does not capture a society’s tendencies towards stability and harmony’ (Institute for Economics and Peace, 2018, April 1st, 2021). Positive peace on the other hand is broader than the absence of war or violence. It focusses on ‘the resilience of a society, or its ability to absorb shocks without falling or relapsing into conflict’ (Institute for Economics and Peace, 2018, April 1st, 2021). As stated by Paffenholz, Abu, And & McCandless (2005, p. 3), positive peace is ‘the presence of social justice’. Peacebuilding interventions based on positive peace aim at addressing the more in-depth processes that stimulate and maintain societal change. For a society to not relapse into conflict and address in-depth processes, it is important to get to know the core cause(s) of the conflict and the involved local, national and/or international actors. Examples of cause(s) are a lack of natural resources, political oppression, or ethnic grievances. To create sustainable peace, it is important to understand what different layers and dynamics must be targeted to realize positive peace. Peacebuilding interventions based on positive peace are quite context dependent and therefore can be very different in nature. (Monnard & Sriramesh, 2020; Scharbatke-Church, 2011)

Peacebuilding evaluation

According to Çuhadar-Gürkaynak et al. (2009, p. 288) ‘evaluation is a complex methodological endeavor, especially if the goal is impact assessment. An intervention’s impact is determined by examining the larger changes initiated by the intervention within the general context [...]’. When focussed on impact of an intervention within evaluation, there is an emphasis on context which already seems to point towards more of a positive notion of peace. This notion of positive peace also seems to be present in Scharbatke-Church’s (2011, p. 460) definition of peacebuilding evaluation:

‘the use of social science data collection methods (including participatory processes) to investigate the quality and value of programming that addresses the core driving factors and actors of violent conflict or supports the driving factors and actors of peace (Church 2008).’

The Peacebuilding Evaluation Consortium (n.d.) argues that there are multiple ways one could carry out a peacebuilding evaluation. For example, through internal or external parties or by focussing on all aspects of the intervention or only reflecting on the intervention with some ‘project stakeholders’ (Peacebuilding Evaluation Consortium, n.d., p. 8). It shows that peacebuilding evaluation is not a quite straightforward practice. This also relates to the fact that peacebuilding evaluations can have multiple functions. As stated by Çuhadar-Gürkaynak et al. (2009), peacebuilding evaluation can have five purposes: (1) providing intervening parties and stakeholders with data to advance the intervention; (2) strengthening the accountability of the intervening party to its donors; (3) providing reflection and learning; (4) enhancing the general knowledge in the peacebuilding field; and (5) contributing to theories on the causes and dynamics of conflict (Çuhadar-Gürkaynak et al., 2009).

Although peacebuilding evaluation can have various goals and purposes and is therefore not a clear-cut practice, there does exist an internationally recognized framework that can be used to carry out peacebuilding evaluation. This framework stems from the OECD (the Organisation for Economic Co-operation and Development) and suggests *relevance, coherence, effectiveness, efficiency, impact, and sustainability* as the evaluation criteria for: ‘[...] development interventions, humanitarian aid,

peacebuilding, climate mitigation and adaptation, normative work, and non-sovereign operations,’ (OECD, 2019, p. 5). These six criteria are explained more in-depth in *Table 1*. These criteria, coherence excluded, were first presented in the ‘1991 OECD DAC Principles for Evaluation of Development Assistance’ (OECD, 2019, p. 2). The coherence criterium was added after the revision of the criteria in 2019, ‘[...] to better capture linkages, systems thinking, partnership dynamics, and complexity,’ according to the OECD (2019, p. 3). There does not seem to be criticism present on this adjusted framework in the academic literature.

Revised Evaluation Criteria Definitions and Principles for Use December 2019	
<i>Relevance</i>	The extent to which the intervention objectives and design respond to beneficiaries’, global, country, and partner/institution needs, policies, and priorities, and continue to do so if circumstances change.
<i>Coherence</i>	The compatibility of the intervention with other interventions in a country, sector, or institution.
<i>Effectiveness</i>	The extent to which the intervention achieved, or is expected to achieve, its objectives, and its results, including any differential results across groups.
<i>Efficiency</i>	The extent to which the intervention delivers, or is likely to deliver, results in an economic and timely way.
<i>Impact</i>	The extent to which the intervention has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects
<i>Sustainability</i>	The extent to which the net benefits of the intervention continue or are likely to continue.

Table 1 Revised Evaluation Criteria Definitions and Principles for Use (OECD, 2019, pp. 7-12)

When trying to capture effective peacebuilding evaluation these criteria can (1) increase understanding on what effective peacebuilding evaluation should entail according to the OECD and (2) serve as a guide when developing and executing an evaluation. However, it does not clearly touch upon how to gather the data to be able to evaluate these criteria.

Research methods

Both the different notions of peacebuilding as well as the context dependent character of peacebuilding interventions mentioned previously, can show why evaluating interventions is complex. For example, when evaluating peacebuilding interventions that focus on negative peace, one could use quantitative methods to determine an increase in the number of demobilized militia or a decrease in conflict deaths or number of conflicts, since quantitative research methods capture numbers as shown in *Table 2* of Hennink, Hutter and Bailey (2011). When evaluating peacebuilding interventions that try to realize more in-depth contextual change, hence positive peace, it can be difficult to measure effects in numbers. It may be more easily captured through qualitative research methods since *Table 2* shows that qualitative research methods are better able to capture understanding of changes within society. However, it is important to note that it does not have to be so black and white. A mix of qualitative and quantitative methods can also capture societal change by both trying to understand as well as trying to measure societal change at the same time. It seems that different peacebuilding evaluations have different notions of peace, purposes, and criteria. This does not point towards a particular research method that is most suitable to capture effect.

	Qualitative research	Quantitative research
<i>Objective</i>	To gain a contextualized understanding of behaviours, beliefs, motivation	To quantify data and extrapolate results to a broader population
<i>Purpose</i>	To understand why? How? What is the process? What are the influences or context?	To measure, count, or quantify a problem. To answer: How much? How often? What proportion? Which variables are correlated?
<i>Data</i>	Data are words (called textual data)	Data are numbers (called statistical data)
<i>Study population</i>	Small number of participants; selected purposively (non-probability sampling) Referred to as participants or interviewees	Large sample size of representative cases Referred to as respondents or subjects
<i>Data collection methods</i>	In-depth interviews, observation, group discussions	Population surveys, opinion polls, exit interviews
<i>Analysis</i>	Analysis is interpretive	Analysis is statistical
<i>Outcome</i>	To develop an initial understanding, to identify and explain behaviour, beliefs, or actions	To identify prevalence, averages, and patterns in data. To generalize to a broader population.

Table 2 Key differences between qualitative and quantitative research (Hennink et al., 2011, pp. 16-17)

Research methods and the attribution gap

When trying to resolve the attribution gap, this could be debated. As stated in the *Introduction*, to attribute an intervention to a societal change or impact, it is a necessity to understand the societal context to make such a claim (Denskus, 2012; Development Assistance Committee, 2007). Denskus (2010, p. 149) elaborates further on his argument:

‘Engaging with the complexities of war zones and transitions from an anthropological or qualitative research perspective is not just a means of storytelling or to present anecdotal ‘evidence’ from ‘the field’. It is the tip of the iceberg of how war and peace are unperceived and that has a strong and direct impact on how and why we carry out evaluation in such environments the way many organizations and professionals do.’

He argues that quantitative research methods do not capture the complexity properly, since it ‘[...] follows a strict development and peacebuilding discourse that favours measurable outputs, glossy reports and treating the evaluation process as a ‘non-activity’ – an activity that is often not supposed to engage participants or raise fundamental questions,’ (Denskus, 2010, p. 151). Based on *Table 2*, this statement seems to be convincing. This is also reflected by Thomas (2005), who states that qualitative research methods are more capable of capturing the causal complexity within Political Science than quantitative research methods. Thomas (2005, p. 862) explains: ‘If the world is characterized by complex multi-causality, isolating individual variables and testing them across time will not capture how they interact; indeed, it is inconsistent with the ontology of complex causality.’ This is because quantitative research mostly focusses on measuring one variable and its effect on Y, while qualitative methods capture the causal mechanism. It is argued by Thomas (2005, p. 861) that qualitative research methods are ‘[...] more representative of the social world than conventional statistical inference.’

Although qualitative methods are argued to be better able to capture complexity and causal mechanisms and therefore seem to attribute an intervention to particular changes or impacts within society best, Bamberger (2012) argues that qualitative methods also come with limitations within peacebuilding evaluation. He argues that the qualitative findings generally show the impact of the whole intervention instead of clearly measuring ‘[...] specific contribution of different components or approaches of the program,’ (Bamberger, 2012, p. 4). He furthermore states that qualitative methods are so strongly related to the context that they are difficult to generalize. Additionally, he finds that qualitative findings can be heavily affected by the subjectivity of the evaluator. Since Bamberger (2012)

also recognizes the limitations of quantitative methods within evaluation, he argues for a mix of both methods to make peacebuilding evaluation more effective. According to Bamberger (2012, p. 4) quantitative and qualitative research methods can complement each other because they, amongst other reasons, enhance '[...] the validity or credibility of evaluation findings by comparing information obtained from different methods of data collection [...]'. When 'equal weight is given to QUANT and QUAL approaches' (Bamberger, 2012, p. 11), he describes the use of methods as follows (Bamberger, 2012, p. 11):

'QUANT surveys are combined with a range of different QUAL techniques. Sometimes the latter focus on the process and contextual analysis, in other cases the focus is on the same unit of analysis as the surveys (e.g., individuals, households, communities, organizations) but different data collection methods are used.'

These different qualitative methods mentioned by Bamberger (2012, p. 16) are: 'in-depth interviews, key informants, participant observation, non-participant observation, case studies, client exit interviews, simulated patient studies, video or audio recording, photography, document analysis, artifacts, group interviews (e.g., focus groups, community meetings), participatory group techniques [...], internet surveys.' When focussing on the attribution gap and hence the need to understand societal context, (some of) these qualitative data collection methods seem to be a necessity within peacebuilding evaluation.

In sum, peacebuilding itself is a broad concept which is highly dependent upon the core cause(s) of the violent conflict and/or the notions of peace. Peacebuilding interventions can therefore be quite different in nature. Peacebuilding evaluations furthermore exist out of diverse methodological practices and have different purposes or actors executing it. There is an internationally acknowledged evaluation framework present to guide interveners through their evaluation. In general, there does not seem to be a specific preference for a particular research method within peacebuilding evaluation. However, when narrowing the scope to the attribution gap, the importance of capturing context and gain understanding seems to stand out, which is best captured by qualitative research methods. Therefore, qualitative research methods seem to be a necessity when trying to resolve the attribution gap within peacebuilding evaluation. To continue this debate, this thesis sheds light on the attitudes, perspectives, and experiences of the several contributing organizations on effective peacebuilding evaluation with a focus on the attribution gap and the use of quantitative and qualitative research methods within their peacebuilding evaluations.

Methodology

This chapter first operationalizes the concepts of the research question already introduced in the previous chapters. Then, it explains why a deductive qualitative method was chosen for this thesis. Afterwards, it will elaborate on the path towards the method of data collection and touches upon the method of analysis, the validity and reliability of the thesis.

Operationalization

Within the research question, effective peacebuilding evaluation, the attribution gap and research methods are the concepts that are investigated. These concepts are explained in the previous chapters. To make them more measurable these concepts are operationalized below.

Effective peacebuilding evaluation

The first concept to be operationalized is effective peacebuilding evaluation. To measure what effective peacebuilding evaluation means to the contributing organizations, the research focusses on what the criteria of effective peacebuilding evaluation are according to the organizations. It does not only try to understand what these criteria are, but also if these are similar to the internationally recognized OECD criteria for peacebuilding evaluation. Understanding more about what effective peacebuilding evaluation is to the organizations, explains what organizations see as important and further choices that are made when executing the evaluation. For example, when peacebuilding evaluation is perceived as effective when the local population is included in the execution of the evaluation, the choice for certain research methods within the evaluation could be dependent upon the expertise of these local people. Moreover, getting to know more about the similarities or differences between the organizations' criteria and the OECD criteria, gives insight into the use of the internationally recognized evaluation framework in practice and if this framework is perceived as useful.

Attribution gap

The attribution gap is a concept that can be experienced by organizations and something they show a particular attitude towards. The ways in which experiences with the attribution gap can be measured is by examining if the attribution gap is experienced at all and how it is experienced. The organizations could experience the attribution gap because of different reasons, like the nature of the conflict being related to a lot of socio-economic, political, and ethnic factors which makes it difficult to attribute a peacebuilding intervention to (a) societal change(s) or impact(s). Another reason why the attribution gap could be experienced, could be due to the lack of capacity to carry out comprehensive evaluations. Getting to know more about the experiences with the attribution gap, increases understanding on how frequently this gap is experienced and possibly what causes this gap according to the contributing organizations. Additionally, ways in which attitudes towards the attribution gap can be measured is by analysing if: (1) organizations state that the attribution gap should be prioritized and/or resolved within peacebuilding evaluation; or (2) if organizations do not pay a lot of attention to the attribution gap due to a possible lack of priority and/or capacity. When understanding attitudes towards the attribution gap, more knowledge will be gathered around the way the attribution gap is perceived and dealt with. Without knowing how the attribution gap is perceived by the contributing organizations, it does not seem to be relevant to understand what research methods are preferred to face this gap.

Research methods

As stated in the previous chapters research methods can be divided into quantitative and qualitative methods. These methods are measured by capturing: (1) attitudes towards quantitative and qualitative research methods; and (2) the use of both methods within peacebuilding evaluation that attribute an intervention to (a) societal change (s) or impact(s).

In order to measure attitudes towards these research methods the research focusses on: (1) the way quantitative/qualitative research methods are viewed; (2) the extent to which quantitative/qualitative research methods are preferred over qualitative/quantitative research methods;

(3) the ways in which quantitative/qualitative research methods are argued to resolve the attribution gap; and (4) the ways in which quantitative/qualitative research methods are argued to fulfil the criteria of effective peacebuilding evaluation. By understanding these attitudes, it becomes clear what the contributing organizations think about the two research methods within evaluation and which one they (possibly) prefer over the other to face the attribution gap. In addition, these attitudes capture if the organizations argue that they are already able to resolve the gap and execute effective peacebuilding organization.

Thereafter, the use of these different methods can be measured by analysing if the goal of the research method is to 'measure, count or quantify' (quantitative) or understand the context, the process and/or the influences (qualitative) (Hennink et al., 2011, pp. 16-17). How data is measured through these methods can be captured by focussing on the use of 'population surveys, opinion polls, exit interviews' (quantitative) or 'in-depth interviews, observation, group discussions' (qualitative) (Hennink et al., 2011, pp. 16-17). By measuring the use of quantitative and qualitative methods, it becomes more clear what these (preferred) evaluative methods look like to the contributing organizations.

Methodological approach: Qualitative

The goal of this research is to gain understanding on how peacebuilding evaluation could become more effective, through understanding what research methods are perceived as more satisfactory to measure the impact of a peacebuilding intervention. Therefore, the methodological approach of this thesis is a qualitative approach. This approach is deductive since the gathered data on attitudes, perspectives, and experiences contribute to the theoretical debates present in the literature. The gathered data tests the theoretical debates that form the starting point of this research.

The process towards a method of data collection: Semi-structured interviews

To contribute to the debates surrounding research methods within peacebuilding evaluation and the attribution gap, this thesis first intended to do an exploratory Most Similar, Different Outcome (MSDO) case study on a peacebuilding organization and an evaluation organization. Such a case study focusses on cases with similar circumstances and different outcomes to understand what causes variation in an outcome (Gerring, 2017). Through such a data collection approach, this thesis intended to understand how the evaluations (the outcomes) of the two similar organizations could differ, while the organizations evaluated similar interventions. It was argued that this could be explained by the use of different research methods within the evaluations by these organizations. However, this approach came with a lot of limitations. First of all, because these two organizations were too different in nature to compare through a MSDO approach. They therefore were not able to be selected for such a case study. Additionally, if these organizations could fit the MSDO approach, it would have been very difficult to find a similar intervention that both organizations evaluated since peacebuilding interventions are so context specific. Moreover, at that time, executing these in-depth interviews seemed not plausible within the time frame of this thesis.

Therefore, it seemed more feasible to analyse peacebuilding evaluations of different peacebuilding organizations. Comparing the trends visible in the different evaluations could show the development of the use of qualitative and/or quantitative research methods and provide information around the attribution gap. However, this analysis also faced obstacles. First, to make the study comparable, the evaluations that had to be gathered had to focus on similar contexts, time frames and interventions. To tackle this problem, two peacebuilding evaluations could be requested from different organizations, a contemporary one and one from a decade ago. However, the second obstacle arose. Detailed peacebuilding evaluations were not that easy to find. There did not seem to be a clear solution to this problem. Gathering similar detailed evaluations became difficult, especially within the time frame of this thesis.

Eventually, also related to the time left to gather data, it became clear that it was needed to just try to get into contact with various organizations that carry out and/or evaluate peacebuilding interventions to see if there were possibilities to gather data in some way. After posting a post on

LinkedIn, contacting a network of people with contacts in this field and around seventeen different organizations or people that were somehow related to peacebuilding interventions; eventually six organizations were interested to contribute to this thesis. These organizations were willing to have a conversation regarding their attitudes, perspectives, and experiences on peacebuilding evaluation. The method of data collection therefore returned to the initial idea of gathering data through interviews. After all, although the schedule was tight, it was possible to carry out six interviews within a time frame of three weeks. With regards to the sampling method of this thesis, the contributing organizations were thus found and contacted through their own websites, the network of the thesis author or were recommended by and/or responded themselves at the LinkedIn-post. In other words, the participants were selected through three non-probability sampling methods: purposive sampling (selecting most useful cases), convenience sampling (selecting available cases), and voluntary response sampling (selecting cases that choose to volunteer). With these sampling methods comes bias and a lack of external validity that is elaborated further on in this chapter.

Descriptive case study

The qualitative method of data collection is a case study. As explained by Gerring (2017, p. 28): ‘an intensive study of a single case or a small number of cases which draws on observational data and promises to shed light on a larger population of cases.’ Since a case study does not intervene, control, or manipulate, it is of explanatory nature. (Gerring, 2017)

As stated in the *Introduction*, the six organizations are different in nature and remain anonymous for various reasons. However, to understand the diversity of these actors, *Table 3* provides an overview of their general characteristics and definitions of peacebuilding. Organization 3, 4 and 5 are relatively large organizations. They for example have international staff in the intervention areas and develop the humanitarian, development, and/or peacebuilding programs. They sometimes have contracts with local organizations to carry out the interventions that are part of the program. The other two non-governmental organizations, organization 1 and 2, are relatively small organizations in comparison to the previously mentioned organizations. In practice, this means that they are mainly contracted as the local organizations. Therefore, they stand in close contact with the people from the intervention area and support these local people with executing the intervention. Organization 6 mostly carries out evaluations for the Dutch government and thus does not carry out interventions itself. Although these organizations are quite different, they all have something to do with peacebuilding interventions and evaluations. Therefore, this case study is a descriptive case study ‘[...] to capture the diversity of [...]’ attitudes, perspectives, and experiences on effective peacebuilding evaluation, with a focus on the attribution gap and the use of quantitative and qualitative research methods within peacebuilding evaluation (Gerring, 2017, p. 58).

Within this descriptive case study, the method of data collection is semi-structured interviews. This means that the structure of the interview is set, like certain topics or questions, but there is still some room for open questions (Schmidt, 2004). This fits the research since it is important to have a structured outline to steer the interviews towards the research topics of this thesis and gives at the same time room to ask questions on other topics they mention and/or to clarify the perspectives and approaches of the participants. The interview protocol that was used during the interviews can be found in *Appendix I*. The questions of this protocol are based on the theoretical debates and the research questions that followed from these debates. Since four out of six interviews were preferred in Dutch, the interview protocol was translated to Dutch (see *Appendix I*). All in all, six semi-structured interviews took place. One per organization. Five interviews were carried out with one person of the organization. One interview with two participants. To guarantee their anonymity, their positions within their organizations cannot be explained. However, they all do work at the head office of their organizations, and some have a decade or more of experience in the peacebuilding field. Additionally, some of the participants were chosen within the organization as best candidates to participate in these interviews, which could say something about their expertise. All seven participants gave consent to record these

interviews, under the conditions that the data was erased after transcription, and all data was made anonymous. Afterwards, these interviews were transcribed.

Contributing organizations	Type of organization	Central theme within interventions	Definition of peacebuilding interventions	Notion of peace	Mostly contracted as
<i>Organization 1</i>	Non-governmental organization	Peacebuilding	Practices that support safe public societal debates	Positive peace	Local partner
<i>Organization 2</i>	Non-governmental organization	Peacebuilding	Practices that stimulate individual talent development	Positive peace	Local partner
<i>Organization 3</i>	International non-governmental organization	Humanitarian aid and development (peacebuilding is part of these intervention)	Practices that support local voices in realizing security, justice, and preferred governmental outcomes	Positive peace	International partner
<i>Organization 4</i>	International non-governmental organization	Humanitarian aid and development (peacebuilding is part of these intervention)	Practices that stimulate resilience and opportunities to voice opinions	Positive peace	International partner
<i>Organization 5</i>	International non-governmental organization	Peacebuilding	Practices that contribute to human security and peaceful and just societies	Positive peace	International partner
<i>Organization 6</i>	Evaluation organization	<i>None</i>	Practices that target the core drivers of a conflict	Positive peace	Evaluating partner

Table 3 Characteristics of the contributing organizations

Methods of analysis: Discourse analysis

According to Smeets (2018), the two most frequently used methods of analysis within Political Science are content analysis and discourse analysis. Hardy, Harley and Philips (2004, p. 20) explain that '[...] discourse analysis focuses on the relation between text and context' while '[...] content analysis focuses on the text abstracted from its contexts'. To understand the attitudes, perspectives, and experiences of the contributing organizations, the appropriate method of analysis within this thesis is therefore a discourse analysis. Through a discourse analysis, this thesis can gain understanding of the dynamics behind effective peacebuilding evaluation. A content analysis on the other hand seems less appropriate, since it '[...] focuses on the message that is communicated by the text, whether that is intentional or unintentional, explicit or implicit,' according to Smeets (2018, slide 6). This thesis is not necessarily interested in just investigating the interviews texts itself, but rather wants to relate these texts to the peacebuilding evaluation context (Hardy et al., 2004). (Smeets, 2018)

This discourse analysis mostly tilts towards a deductive coding method because most coding themes used during the coding process are built upon the theoretical debates presented earlier. The coding scheme used during the coding process can be found in *Appendix II*. Due to the semi-structured character of the interviews, unforeseen topics arose during the gathering of data. Additionally, the participants were asked if they wanted to touch upon any remaining insights or obstacles. Since this data could present interesting information regarding effective peacebuilding evaluation, this data was coded under the variable *other input regarding to effective peacebuilding evaluation* and analysed through an inductive method of open coding. Afterwards these codes were made into categories through axial

coding and selective coding. This process eventually led to certain variables and related indicators which are also presented in *Appendix II*. Lastly, all coding was done through Atlas.ti 8.

Validity

Internal validity

Based on the literature touched upon earlier, this thesis tries to test if the attribution gap decreases when the amount of qualitative research methods increases within peacebuilding evaluation. When studying the attitudes, perspectives, and experiences of various organizations, this relationship can be tested, in a non-statistical way. However, it is difficult to rule out other external factors that could influence this relationship. Examples are the way the qualitative methods are carried out or the nature of the violent conflict which could make resolving the attribution gap more difficult or more easy. Therefore, the internal validity of this research is not rock solid. Nonetheless, gathering and describing these attitudes, perspectives, and experiences does increase understanding on the way this relationship is experienced first-hand.

External validity

The external validity of the case study seems to be quite low, since a small number of organizations only show a fracture of attitudes, perspectives, and experiences on effective peacebuilding evaluation, the attribution gap and research methods of the population (the population being all organizations executing and/or evaluating peacebuilding interventions). This makes it difficult to generalize the findings. Nonetheless, it is possible to generalize less context-specific findings of this research which are touched upon in the *Discussion*. Furthermore, the non-probability sampling methods make the cases less representative for the total population which lowers the external validity of the research. Convenience sampling for example does not necessarily selects cases that are representative for the population. Purposive sampling and voluntary response sampling both lead to bias which also affects the generalizability of the research. The selection of cases through purposive sampling is based on the knowledge and experience of the researcher regarding the most appropriate cases for this research. However, such a perspective is always subjective and therefore biased. Voluntary response sampling also brings bias because the organizations that volunteer to participate might do so out of particular interest(s) or other intrinsic motivations.

Reliability

What might have impacted the reliability in a positive way, is the guarantee that all data remains anonymous. This could have made participants less reluctant in their answers, which makes the data consistent. However, a discourse analysis always comes with bias due to the subjectivity of the researcher carrying out the discourse analysis (Smeets, 2018). If a data analysis is carried out by someone else, it can lead to different outcomes and therefore affect the consistency of the data. Moreover, when carrying out the interviews, interview experience increased. With time it became easier to get the participants back on track when they lost sight of the interview questions. Therefore, this research suffers from experimenter bias. The behaviour of the interviewer was not consistent throughout the data collection, which impacts the consistency of the data. Another point that lowers the consistency of data collection, and therefore the reliability, are the requests of two out six organizations to have a look at the interview protocol before the interview. To prevent losing these participants, the protocol was sent beforehand. The interviews that were already carried out did not have this advantage. This could have impacted the way the participants responded to the interview questions and therefore could provide very different data than when the questions were asked without reading the protocol beforehand.

Chapter 1: What criteria do the contributing organizations use to assess effective peacebuilding evaluation?

Before diving into the organizations' take on the attribution gap and research methods, this chapter first explores what effective peacebuilding evaluation means to the contributing organizations by capturing their criteria of effective peacebuilding evaluation. Their criteria could influence their experiences with and attitudes towards the attribution gap and/or their take on the use of certain research methods within the evaluation as explained earlier. In addition, it tries to understand if these criteria meet the OECD criteria and if these are perceived as useful. Therefore, this chapter first provides some background information on these criteria, before diving into the perspectives of the organizations.

Criteria according to OECD

As presented in the *Theoretical Framework*, the OECD tries to provide guidance through their six criteria for effective peacebuilding evaluation (OECD, 2019). As mentioned before these criteria were revised in 2019. The OECD (2019, p. 2) describes that during the revision process:

'The consultations found widespread support for and use of the criteria. Respondents clearly preferred to maintain the main body of the current set of evaluation criteria, recognising their universal acceptance and usefulness. Many respondents highlighted the value of the criteria in bringing standardisation and consistency to the evaluation profession and evaluative practice.'

To paint a clear picture of this consultation, it consisted of 691 survey responses of which the largest share came from independent consultants (around 145 surveys), international non-governmental organizations (around 97 surveys), private actor companies (around 62 surveys), central government/academic institutions (around 55 surveys), and research organizations/think tanks (around 48 surveys). The other surveys came from various sources, like the UN Agency HQ, foundations, 'other non-governmental/civil society...', local governments, and 'other inter-governmental institutions' (OECD, 2018, p. 3). In other words, this revision consisted of quite a diverse amount of feedback. The obstacles that were pointed out are (OECD, 2019, p. 3):

'At the same time, requests were received for clarifications of certain concepts. Many pointed to challenges with the way the criteria are applied in practice. Particularly problematic is the tendency to cover too many criteria and questions. While the Quality Standards for Development Evaluation are clear that the use of all the criteria is not mandatory, and that other criteria may be used, in practice they can end up being applied mechanistically without sufficient consideration of the evaluation context and intended purpose. There was also concern that the original set of criteria did not adequately encompass the 2030 Agenda narrative and current policy priorities. Some felt the criteria were too project-focused and did not sufficiently address issues such as complexity and trade-offs, equity, and integration of human rights and gender equality. Many requested enhanced guidance on implementation of the evaluation criteria, to improve their use and to contribute to enhanced evaluation quality.'

According to the OECD, the revised criteria respond to these obstacles and are thus rooted in a deliberate process of consultations regarding the criteria for effective peacebuilding evaluation. Since not a lot of criticism can be found regarding these criteria after the revision of 2019, it can be perceived as a useful framework of which its content is considered deliberately.

However, although the revision is rooted in input of various organizations that go beyond the OECD's members (OECD, 2018), it is important to emphasize possible interests. The OECD consists of 38 country members (listed below) and has four close key partners: Brazil, China, India, and Indonesia (OECD, n.d.). 85 out of the 691 survey responses came from its members (OECD, 2018). Additionally, the countries where most conflict and probably most peacebuilding interventions occur, Middle Eastern and African countries (IISS, 2021), are not members of the OECD (except for Israel and Turkey). All in all, it is important to keep in mind that the revised peacebuilding evaluation recommendations of the OECD are a product of an intergovernmental organization that could be

influenced by the interests of its members, of which most do not experience peacebuilding interventions in their own countries and are mostly related to interventions by funding them. These members could therefore be more interested in designing criteria that show what their funding achieved in order to remain accountable to its own citizens, rather than designing evaluation criteria that provide the executing organizations a learning experience or are e.g. focussed on the inclusion of local voices. (OECD, n.d., June 8th, 2021)

• Australia	• Finland	• Korea	• Slovak republic
• Austria	• France	• Latvia	• Slovenia
• Belgium	• Germany	• Lithuania	• Spain
• Canada	• Greece	• Luxembourg	• Sweden
• Chile	• Hungary	• Mexico	• Switzerland
• Colombia	• Iceland	• Netherlands	• Turkey
• Costa Rica	• Ireland	• New Zealand	• United Kingdom
• Czech Republic	• Israel	• Norway	• United States
• Denmark	• Italy	• Poland	
• Estonia	• Japan	• Portugal	

Table 4 Members of the OECD (OECD, n.d., June 8th, 2021)

Criteria according to contributing organizations

Within this research, the data shows that the contributing organizations vary tremendously in their perspective upon the criteria of effective peacebuilding evaluation. Some do not seem to have an answer, while others do. Some mention the OECD criteria and others do not.

Participant 1 for example did not mention any criteria for effective peacebuilding evaluation (Interview 1). When asking participant 2: ‘what are the criteria of effective peacebuilding evaluation?’ The participant answered with: ‘Yes, only if we knew,’ (translated from Dutch to English) (Interview 2). This already shows a lack of knowledge regarding this topic.

Two other organizations did have an answer to the question and formulated their own criteria for effective peacebuilding evaluation. Participant 3 defined two criteria, which seem to relate to its aims of evaluation. The first criterion relates to *the extent to which the local population is included in the peacebuilding evaluation*. The second criterion of effective peacebuilding evaluation relates to *the extent to which the peacebuilding evaluation makes the evaluator and the local population learn something* (Interview 3). This participant also touched upon ‘the criteria of the IOB’. The IOB (Policy and Operations Evaluation Department) refers to a body that executes independent evaluations for the Ministry of Foreign Affairs of the Dutch government (Ministry of Foreign Affairs, n.d.). Participant 3 argues that the criteria of the IOB are too technical and too much focussed on effectivity and efficiency, which are argued to not suit peacebuilding interventions since peacebuilding processes are transformation processes that are too difficult to compare. The participant would like to see the IOB being more focussed on the second criterion that (s)he proposed. Participant 4 did not mention anything about the OECD criteria and proposed one criterion for effective peacebuilding evaluation which relates to *the extent to which a conflict sensitive approach is integrated in the peacebuilding evaluation* (Interview 4). The participant explained a conflict sensitive approach as follows:

‘conflict sensitive in the sense of good understanding/analysis of the context of the interaction between context and a project and act upon that understanding to avoid negative impact, do no harm, and go one step further; contribute to positive change and promote peace where possible on these.’

Additionally, participant 4 also expressed an emphasis on learning. This participant views this as a critical part of peacebuilding evaluation. However, the participant did not specifically mention learning as a criterion of peacebuilding evaluation. After all, the three criteria put forward by participant 3 and participant 4 differ a lot from the OECD criteria, which shows that their own priorities and aims influence the way peacebuilding evaluation is perceived as effective.

The remaining two organizations do use (some of) the criteria of the OECD. Participant 5 and participant 6 (both from Interview 5) first expressed that they do not have ‘a strong evaluation tradition’ (translated from Dutch to English). They are mostly focussed on monitoring their interventions and are not quite interested in carrying out evaluations that do not provide them a learning experience. However, after this statement, participant 6 did mention that they use the OECD evaluation criteria, of which the *sustainability* criterion is the most important to them. They are interested in *the extent to which the realized changes in actors remain present*. Lastly, participant 7 touched upon both the OECD criteria and their own quality criteria (Interview 6). They are interested in *the extent to which the OECD criteria and their own quality criteria are met during an evaluation*. Participant 7 showed a critical attitude towards the OECD criteria. In the opinion of the participant these should not be called criteria. The participant did not explain this train of thought since it was argued to be too comprehensive for this interview. The quality criteria try to guarantee the quality of the peacebuilding evaluation, by explaining how to carry out a research. Later in the interview, participant 7 also highlighted the importance of learning during peacebuilding evaluation. However, the participant does not think it is wise to focus mostly on learning: ‘If you do not know if you are effective [...] how do you know if you are going the right direction?’ (translated from Dutch to English). When this participant was asked if the inclusion of the local population could better peacebuilding evaluation, the participant argued that it could be relevant since it could make an evaluation less based on a fictional reality. This participant also argued why peacebuilding evaluation is not that effective: (1) some organizations are not critical on their own interventions because of their belief in their own intervention and knowledge regarding the intervention area; (2) there are a lack of resources; and/or (3) the financial interests in wanting to have a good track record to bring in new funding opportunities also (negatively) impacts the way peacebuilding is carried out. According to this participant, peacebuilding evaluation still has a long way to go in general based on the Evidence for peacebuilding: An evidence gap map (Cameron et al., 2015).

Ten years ago

To understand more of the knowledge of the participants on peacebuilding evaluation criteria, the question was asked if these criteria differ a lot from ten years ago. Again, the data differs a lot. Participant 1, participant 2 and participant 3 did not have an answer to this question. Mostly because they did not work in the peacebuilding field around that time or simply lack the knowledge. Participant 4 did have an answer to the question and argued that a decade ago a conflict sensitive approach within peacebuilding evaluation was less present. One of the participants from Interview 5 mentioned that ten years ago (s)he was the first monitoring and evaluation advisor within the organization. This already shows that evaluation of peacebuilding intervention is a relatively new practice for some organizations. Additionally, the participant explained that the organization did carry out evaluations that were checked on quality by the IOB. The participant was however not able to mention specific criteria and explained that (s)he does not know how these evaluations were carried out. Only participant 7 touched upon the revision of the OECD criteria.

Conclusion

To answer the question of this chapter, it first depends on if an organization can mention criteria. If they do, the following indicators relate to the criteria they use to assess effective peacebuilding evaluation: (1) *the extent to which the local population is included in the peacebuilding evaluation*; (2) *the extent to which the peacebuilding evaluation makes the evaluator and the local population learn something*; (3) *the extent to which a conflict sensitive approach is integrated in the peacebuilding evaluation*; (4) *the extent to which the realized changes in actors remain present*; and (5) *the extent to which the OECD criteria and their own quality criteria are met during an evaluation*. Hence only two organizations use criteria that are linked to the OECD criteria to assess effective peacebuilding evaluation. All in all, the amount of knowledge around the topic differs but is mostly not that comprehensive. Additionally, there does not seem to be coherence in the criteria to assess effective peacebuilding evaluation, but it could

be stated that these criteria rather relate to evaluation aims, like inclusiveness and ownership, than accurately measuring effects.

Relating results to the debates

When relating these results to the debate, this chapter provides understanding on what effective peacebuilding evaluation means. It also shows that peacebuilding evaluation is not always a necessity and therefore self-evident. The criteria that define effective peacebuilding evaluation do not only differ a lot from the criteria of the OECD, but they are also mostly criticized. There does not seem to be a 'widespread support for and use of the criteria' or a recognition of the 'universal acceptance and usefulness' of the OECD criteria (OECD, 2019, p. 2). Lastly, since the approaches to effective peacebuilding evaluation are so different, these differences could make organizations learn from each other. They for example could discuss why some point towards local ownership and learning as important elements within effective peacebuilding evaluation.

Chapter 2: In what ways and to what extent do the contributing organizations experience the attribution gap?

Since it is clearer how the contributing organizations view effective peacebuilding evaluation, this chapter examines the experiences and attitudes towards the attribution gap. Do they experience the attribution gap? How do they experience this? What is their take on this problem? This chapter first sheds light on the experiences with the attribution gap, which is followed by the attitudes the contributing organizations show towards this gap.

Experiences

All participants experience an attribution gap, which is not quite surprising since one can assume that the attribution gap will always remain present to a certain extent in a peacebuilding context. Simply because an evaluator will never be able to isolate the intervention from the dynamic and ever-changing context where the intervention is embedded in. As described by participant 3 (translated from Dutch to English):

‘Within an evaluation, you hope for some kind of baseline, but there are so many external factors that are of influence, that it is blurred. You can never say, this can be attributed to the program. Especially in these kind of conflict areas the context changes continuously and that makes it very difficult to draw hard conclusions.’

To gain more understanding on how the attribution gap is experienced, the following section dives into the reality of attributing a peacebuilding effort to societal change(s) or impact(s).

When participant 1 explained more about the experience with the attribution gap, it became clear that attributing a peacebuilding intervention to a change in the behaviour of authorities can only be based on assumptions. Participant 1 executes peacebuilding interventions in, amongst others, Burundi, Mali, China, and DRC Congo. (S)he explained:

‘And, you know, it's it's very difficult to ask authorities in such countries: Is it thanks to us that you did that? We never do that because we would immediately endanger a project. [...] They're not accessible. They don't want to speak with an NGO.’

The political sensitivity of the interventions makes this organization feel obliged to make attribution claims based on assumptions instead of clear confirmations. The following section shows how such a process of assumptions, or in the words of the participant ‘impact journey’, occurs:

‘Another example still in Burundi, we had a campaign. There was this drug, which is called Boost. [...] But it was a taboo. Nobody wanted to speak about that and especially authorities, because maybe some children of authorities were also trapped into that. And the campaign was [x], it's quite simple, but it revealed the phenomenon. And it also pushed the authorities and the families towards more empathy for this. For these young people, they're not criminals. [...] You know, they're sick. And so they work with the local association from former addicts for this campaign. This association was criminalized. It was delegitimized, police were saying, 'we don't work with junkies', 'we don't work with drug addicts'. But thanks to this campaign, they became really a legitimate partner for the authorities to work with. Well, and then there was some announcement from the Ministry of Interior that they're going to reintroduce medication. They're going to open a rehab center. [...] I've got this to show you, that when you work with very vulnerable groups, which without the local platform will have never got this opportunity to be respected and to be useful for their own community. Yeah, I guess it's a sign of peace building, right?’

Participant 2 also expressed difficulties regarding the attribution of an intervention to an impact. (S)he explained that this is due to the nature of their interventions, which focus on stimulating talents within youth across the world. The participant explained that it is very difficult to measure when these talents are stimulated within an individual and how this contributes to the community and the society of this

individual. The participant concluded that it is too difficult for them to measure if their interventions can be attributed to a certain impact.

Besides the political sensitivity and the nature of an intervention that influence the attribution gap, participant 3 addressed the distinction between contribution and attribution. The participant argues (translated from Dutch to English):

‘You try to prove that you have contributed to it. So contribution, and then you still have contribution and attribution, which are two different things. Eh hè, because what you can contribute, you can describe. For example, that we organized a meeting. We brought people to a peace talk, just some examples. If that has attributed to a change, that is way more difficult, hence you must make assumptions. You never know for sure.’

This case demonstrates that it is experienced as more difficult to show the attribution than the contribution of an intervention. Contribution relates to what the intervention did in the intervention area, while attribution relates to the changes that were realized due to the intervention. It furthermore also emphasizes that attributing an intervention to impact has to be based on assumptions.

Participant 4 explained that (s)he finds it hard to isolate the impact of an intervention from the social, political, and economic factors and to deal mostly with ‘an indirect correlation’. The participant argues that these difficulties make it challenging to get to know the negative and positive impacts.

The distinction between contribution and attribution was again emphasized in Interview 5. Participant 6 answered the interview question with: ‘Well, we never talk about attribution anyway,’ (translated from Dutch to English). Why they do not focus on attribution is explained in the next section regarding their attitude towards the attribution gap. Participant 5 furthermore also emphasized political sensitivity. (S)he argued that changes in behaviour that could be attributed to an intervention are mostly not confirmed by authority figures and are therefore always based on assumptions.

Participant 7 mentioned that due to the various number of relevant variables that are embedded in the complex context, (s)he experiences it as quite comprehensive and difficult to attribute an intervention to a change or development. Additionally, this participant emphasized the frequent lack of access to the area of the intervention, which limits the provision of information.

Attitudes

Although it is relevant and interesting to understand if and how the attribution gap is experienced, it does not necessarily mean that the organizations are interested in (partly) resolving this gap. This section sheds light on the attitudes of the participants regarding attributing a peacebuilding effort to societal change(s) or impact(s).

Participant 1 and participant 7 did not express a clear attitude towards the attribution gap. They just mentioned that it is a difficult matter. Participant 2 on the other hand articulated the willingness to measure the impact of their interventions, but (s)he explained that (s)he does not know how to do this exactly. Participant 3 showed a clear attitude and expressed that attribution does not have the highest priority within peacebuilding evaluation. Although this participant did emphasize the importance of being able to describe impact of an intervention, (s)he argued (translated from Dutch to English):

‘If I know that organizations are organized well, thoroughly think about their influence on policy, and the policy changed. Is this 100 percent done by them, or 90 or 50 or maybe only 10 percent? If they developed the capacity to execute it well, then I’m very content.’

According to participant 4, the attribution gap is ‘[...] inherent to peace building evaluations [...]’. Participant 4 viewed it as an ongoing challenge without easy answers and therefore emphasized again that a conflict sensitive approach and the awareness of the attribution gap must be integrated in a peacebuilding evaluation. As mentioned earlier on in this chapter, participant 5 and participant 6 prefer to focus on contribution instead of attribution. Participant 6 argued that it is ‘[...] always about contribution,’ because the peacebuilding field is so complex that it is mostly not possible to claim that changes occur only due to your own intervention(s) (translated from Dutch to English).

Conclusion

To answer the research question of this chapter, the attribution gap is experienced in different ways by all organizations. Important factors that make it difficult to attribute impact to a peace effort are political sensitivity, the nature of the peacebuilding intervention, the indirect correlations embedded within a complex context, the inherent nature of peacebuilding evaluations, and the lack of access to the intervention area. This chapter shows that the attribution gap can have multiple reasons. Additionally, the chapter shows that attribution claims are mostly based on assumptions. These assumptions together form the plausible causal relation. Therefore, it could be argued that attribution claims are mostly based on process-tracing, which is a qualitative research method that investigates ‘[...] the intervening causal process – the causal chain and causal mechanism – between an independent variable (or variables) and the outcome of the dependent variable,’ (Bennet and Checkel, 2014, p. 6).

When diving deeper into the attitudes towards the attribution gap, it becomes clear that attributing an intervention to an impact is not always prioritized. Only one organization clearly stated that it tries to resolve this problem within peacebuilding evaluation. Other organizations that did show a willingness to measure impact of the intervention, mostly shift away from attribution towards contribution since attributing an intervention to impact is perceived to be too difficult or even impossible.

Relating results to the debates

When relating these results to the debate regarding the attribution of a peacebuilding intervention to (a) societal change(s) or impact(s), this chapter demonstrates that the organizations are aware of the attribution gap and consider it problematic. It however shows a contradiction regarding the importance of resolving this gap. While the organizations on the one hand do emphasize the importance of exploring the impact of their intervention, on the other hand, they demonstrate that they do not know how to face the attribution gap and/or perceive resolving this gap as a priority. The latter is not always related to a lack of interest. It also relates to a lack of resources, like time or money, that are needed to measure various external contextual variables to capture attribution. Only one organization specifically addressed the need for a conflict sensitive approach that analyses the interaction between the context and the intervention within evaluation. This seems to be an approach that is in line with the theoretical debate that also emphasizes the importance of capturing context to resolve the attribution gap.

Chapter 3: Do the contributing organizations prefer certain research methods within peacebuilding evaluation to capture the impact of their peacebuilding interventions?

After examining what effective peacebuilding evaluation means and in what ways and to what extent the attribution gap is experienced, this chapter tests if the organizations would argue that qualitative research methods can help to resolve the attribution gap. The chapter first examines the ways in which the organizations perceive and use quantitative and/or qualitative research methods within peacebuilding evaluation to capture impact. Thereafter, it becomes more clear in what ways these research methods are expected to resolve the attribution gap and how these methods fulfil their criteria of effective peacebuilding evaluation.

Attitudes and use

Before diving into the attitudes towards and the use of particular research methods, it is important to emphasize the inability of some participants to systemically explain their evaluative research methods. This could relate to the lack of knowledge and/or priority around peacebuilding evaluation and the attribution gap touched upon in the previous chapters. It might affect the way evaluation methods are carried out. Although some examples and descriptions were given, which are presented below, it does not always become quite clear what their quantitative and/or qualitative practices actually entail. Nonetheless, the following section tries to shed light on these attitudes and uses of these methods by the organizations. Some participants do not prefer qualitative research methods over quantitative research methods, while others do.

Preference for a mix of both

Participant 1 explained that it does not prefer qualitative research methods over quantitative research methods or vice versa and argued that they should complement each other. Mostly because quantitative data in itself can give a skewed view of the impact of an intervention. (S)he explains that tracking the impact of online interventions does require mostly quantitative research methods and is therefore mostly used. However, the participant did show a critical stance towards the use of these quantitative research methods: '[...] but I can tell you, because I've been working in that, it can be such a manipulation.' The participant did not know a lot about the qualitative research methods that are used by the organization within peacebuilding evaluation. Additionally, participant 4 did not prefer a research method over another; but did mention that quantitative research methods are not sufficient to measure outcomes, outputs, and impacts of peacebuilding. Therefore, a mix of both methods is preferred, which must measure the attitudes, perceptions, and changes in behaviour because of the intervention. Surveys and questionnaires are examples of quantitative methods that are used by this organization. Examples of qualitative methods are interviews, focus group discussions, and conversations during intervention activities. Another participant that did not show a preference for a particular research method within peacebuilding evaluation is participant 7. The method that is used, is based on the research question and the theory behind the intervention. This participant did point towards the difficulty of executing an experimental research within a peacebuilding context, which makes it most of the time more likely that qualitative methods are used in practice. It is however emphasized that this is based on the context and the limitations of peacebuilding, not because there is a preference. Examples that are used when needed are quantitative surveys or qualitative interviews. This participant also expressed criticism towards the use of research methods by various non-governmental organizations. (S)he argued that the quality of the evaluations is not always that high and that the qualitative method of outcome harvesting has a high risk for bias. The latter is a qualitative research method that gathers outcomes through group discussions with local partners. The participant argued: 'Actually all the evaluations that have been undertaken with that method, are more useful for, put it bluntly, but a flyer [...]. Then for a critical evaluation,' (translated from Dutch to English).

Preference for qualitative research methods

The participant of Interview 2 however explained that the organization only uses qualitative research methods to measure the impact of an intervention. Qualitative research methods are preferred over quantitative research methods because this organization prefers to measure personal growth through qualitative methods, like interviews. This already shows that the choice for particular methods, depends upon the nature of the intervention. However, (s)he also explained that the organization focusses on output like the number of projects or trainings, which indicates more of a quantitative approach. The participant proposes that other relatively larger INGOs might be better equipped to carry out quantitative methods due to their larger programs. (S)he argues that smaller INGOs could complement these qualities by providing their qualitative research methods expertise. Participant 3 also showed a critical stance towards quantitative research methods, because ‘numbers do not say that much’ about complicated peacebuilding transformation processes (translated from Dutch to English). This stance relates to the interests in measuring beyond output; namely the outcomes and impacts of an intervention that require qualitative methods according to the participant. (S)he does use quantitative methods, which ones are not clear, but as long as it is clarified with qualitative methods. These quantitative methods are also used because this is required by ‘the Ministry’. The participant explained that although this party also ask for qualitative narratives to support the quantitative data, it also wants the organization to put targets on these quantitative numbers. (S)he said: ‘[...] then you lose me completely,’ (translated from Dutch to English). The evaluations that are carried out by the organization of participant 3 mostly focus on outcome harvesting; to ‘tell a story’, capture the different steps, and show the unexpected results of an intervention (translated from Dutch to English). Although participant 5 and participant 6 focus mainly on monitoring instead of evaluation and contribution instead of attribution, participant 5 also argued that it makes the participant ‘unhappy’ when the organization has to put the impact in quantitative numbers for the IOB (translated from Dutch to English). Although (s)he mentioned that these methods can be helpful to track progress, it is argued to be too difficult to define or interpret and therefore does not add a lot of value. Additionally, these participants mentioned that they also mostly use outcome harvesting to gather qualitative data and capture impact within their monitoring processes.

Research methods and the attribution gap

It is important to note that not all participants were specifically asked how their research methods help to resolve the attribution problem, which is elaborated in the *Discussion*. Multiple participants that were asked this question expressed high hopes that their research methods, mostly qualitative outcome harvesting, can capture the contribution of their intervention. However, since they do not focus that much on attribution, these methods do not seem to resolve the attribution gap. Another participant explained that the theory-based approach does seem to be able to (partly) resolve the attribution problem, since it maps the context by including all kinds of contextual factors that influence or are influenced by the intervention. All in all, one participant clearly stated that their methods, that can be both qualitative and quantitative, can help to resolve the attribution gap as long as it is rooted in a theory-based approach.

Research methods and the criteria of effective peacebuilding evaluation

When relating the research methods to the criteria used to assess effective peacebuilding evaluation, again not all participants were asked this question which is explained in the *Discussion* chapter. However, participant 2 argued that they experience a gap between the two. Participant 3 on the other hand argued that its research methods include the local population within the evaluation and lead to a learning experience and therefore fulfil the criteria of this participant. The extent to which the research methods include a conflict sensitive approach depends on the intervention according to participant 4. Participant 4: ‘Has it been standardized? Is everyone doing it yet? No, not yet.’ Finally, the extent to which the research methods of participant 5 and participant 6 fulfil their sustainability criteria is perceived as difficult, due to the difficulty of measuring sustainability. This section shows that it does not seem to be easy for most organizations to fulfil their criteria of effective peacebuilding evaluation with their current research methods.

Conclusion

To answer the main question of this chapter, there is not a clear-cut answer since there is not always a preference for a quantitative or qualitative method. However, all organizations recognize the limitations of quantitative methods within peacebuilding evaluation when trying to measure impact. The organizations that do prefer qualitative methods to measure impact, argue that quantitative methods do not seem to be necessarily sufficient in doing so and mostly only capture outputs, instead of outcomes and impacts. A popular qualitative method is outcome harvesting, which is not always perceived as a research method with high quality. The other half prefers a mix of both methods within peacebuilding evaluation since quantitative methods should always be supported by qualitative methods. It is worth noting that it was not argued that qualitative research methods on its own are not sufficient within peacebuilding evaluation. This chapter furthermore made clear that there is a difference between an attitude towards particular research methods and the actual use of particular research methods within peacebuilding evaluation. This can partly be explained by the guidelines that organizations are required to follow to provide accountability towards their partners. It furthermore does not seem to be that easy to develop research methods that (partly) resolve the attribution problem. Additionally, many organizations find it hard to determine if their research methods fit the criteria that they think define effective peacebuilding evaluation best.

Relating results to the debates

As stated in the Introduction, measuring the effect of a peacebuilding intervention within conflict-affected areas is complex, due to the '[...] multitude of uncontrollable external variables' (Çuhadar-Gürkaynak et al., 2009, p. 295). The solution to this attribution problem was argued to be 'carefully designed evaluative methods,' by Çuhadar-Gürkaynak et al. (2009, p. 296). Since these variables are present in the context of a peacebuilding intervention, qualitative methods were argued to capture these external variables best. The results gathered in this chapter seem to affirm this perspective. Since quantitative methods are argued to not be able to capture context, qualitative methods within peacebuilding evaluation are and therefore seem to be able to help resolve the attribution gap and make peacebuilding evaluation more effective. However, the chapter also shows that it is not only about what methods are perceived as most appropriate to measure impact but also what evaluation methods fit the intervention best. Moreover, when linking these findings to the actual use of research methods within peacebuilding evaluation, some seem to grope in the dark, while others are convinced that outcome harvesting or some kind of process-tracing method are the best way to go about this.

Conclusion

At the start of this thesis, it became clear that the peacebuilding evaluation field stumbles upon various challenges. One of those challenges which limits the effectiveness of peacebuilding evaluation is the attribution gap, which could be faced with ‘carefully designed evaluative methods,’ (Çuhadar-Gürkaynak et al., 2009, p. 296). Since the difficulty of attributing a peacebuilding intervention to (a) societal change(s) or impact(s) mostly relates to the difficulty of capturing the various (changing) external variables within the context of the intervention, the theoretical debates argued that qualitative research methods would be needed within evaluation to capture these contextual variables. This implied that to make peacebuilding evaluation more effective, qualitative research methods within peacebuilding evaluation are better able to (partly) resolve the attribution gap. In other words, when the use of qualitative research methods within peacebuilding evaluation increases, the attribution gap should decrease.

To add to this debate, this descriptive case study tried to understand the different attitudes, perspectives, and experiences of six organizations that execute and/or evaluate peacebuilding interventions regarding effective peacebuilding evaluation, with a focus on how they try to resolve the attribution gap through quantitative and/or qualitative research methods within their evaluations. However, when trying to determine what ‘carefully designed evaluative methods’ mean to these six organizations to resolve this gap, the majority of the participants were not able to systematically explain their evaluation methods, if evaluation was a priority at all. The organizations mostly do not present a clear stance towards effective peacebuilding evaluation and the research methods they use, let alone that they are concerned with the attribution gap and try to resolve this problem. This shows that although the attribution gap is debated in the theoretical debates, in practice the peacebuilding evaluation field does not seem to be there yet.

This thesis tried to answer the following question:

How do organizations that execute and/or evaluate peacebuilding interventions perceive quantitative and/or qualitative research methods within peacebuilding evaluations as more satisfactory in addressing the attribution gap?

First of all, what influences the perspective upon the preference for particular research methods within peacebuilding evaluation are the different perspectives of the organizations upon effective peacebuilding evaluation. The thesis shows that effective peacebuilding evaluation starts with different criteria and hence means different things to different organizations, see *Table 4*. This influences their take on the way that peacebuilding evaluations are carried out, if carried out at all. Secondly, when knowledge around evaluation is lacking and/or it is not that much of a priority (see *Table 4*), the attribution of an intervention to impact is not something that seems to be quite present in the evaluation either. This contrasts the willingness of most organizations to capture the effects of an intervention beyond outputs and measure outcomes and impacts through their evaluations. To capture the impact, it is mostly argued that quantitative methods are not comprehensive enough to fulfil this task in a peacebuilding setting, and therefore qualitative methods are seen as a necessity to measure this or to support quantitative evaluation data. In practice, some organizations use qualitative research methods that capture the contribution of the intervention, like outcome harvesting. Others prefer a mix of both methods as long as it fits the intervention and/or captures the attitudes, perceptions, and changes in behaviour caused by the intervention. Although some data collection methods are provided in this thesis, like interviews or surveys, it does not shed a lot of light on what these qualitative and/or quantitative methods entail, since most participants are not able to explain very clearly how their data is gathered. Concluding, this descriptive case study finds that qualitative research methods are perceived as more satisfactory in addressing the attribution gap than quantitative research methods.

So, what do these findings mean for the academic and societal debates? First of all, it furthers knowledge on what effective peacebuilding evaluation means to different organizations. It shows that the criteria underlying effective peacebuilding evaluation vary a lot within the peacebuilding field and are dependent upon goals of the evaluation and the nature of the intervention. It furthermore adds to the debate on the value and necessity of peacebuilding evaluation and the use of the internationally recognized criteria. It presents that peacebuilding evaluation and the use of the internationally recognized criteria are not always that self-evident as the literature or intergovernmental bodies like the OECD might present. This thesis also furthers debate on the actual importance of addressing the attribution gap, by explaining that many organizations do not show that much willingness to resolve this gap. The results show that the attribution gap is perceived as a challenge, which fits the perspectives of Çuhadar-Gürkaynak et al. (2009), Scharbatke-Church (2011), OECD (2012b), and Anderson and Olson (2003). However, this does not necessarily mean that it is a concern within peacebuilding evaluation. Regarding the academic debate on how to resolve the attribution gap, this thesis adds to the debate by supporting the theoretical theory that qualitative research methods capture the (changing) contextual variables better than quantitative research methods and therefore need to be included within peacebuilding evaluation when trying to attribute a peacebuilding intervention to (a) societal change(s) or impact(s).

The results furthermore show that the peacebuilding evaluation field is not as far ahead in practice as the literature may make it seem, by explaining that the field does not seem to be that concerned with the attribution gap and sometimes even with evaluation. If evaluation tries to capture impact, it sometimes rather focusses on contribution than attribution within peacebuilding evaluation. Other organizations show that they try to measure impact through process-tracing, while a few try to capture impact by using methods that capture the context. In addition, the thesis gives organizations insight into the practices of other organizations. Since the approaches of the organizations are so different from each other, it could inform organizations about the various perspectives, preferences, and options regarding effective peacebuilding evaluation and methodological approaches available and used within their field. It hence could serve as a learning tool. Lastly, the research furthers debate on how peacebuilding evaluation could become more cost-effective by arguing that using only quantitative research methods to capture impact is mostly not beneficial for the quality of the evaluation. If organizations want cost-effective peacebuilding evaluations, quality seems to go over quantity or the two have to go hand in hand.

Contributing organizations	Criteria for effective peacebuilding evaluation	Attribution gap	Research methods
<i>Participant 1</i>	<i>No response</i>	The attribution gap acknowledged, but further attitudes are not present.	No preference for qualitative research methods over quantitative research methods or vice versa. These methods must complement each other because quantitative data can be a manipulation of the actual effect of an intervention.
<i>Participant 2</i>	‘[...] only if we knew.’	The attribution gap is recognized and there is a willingness to measure the impact of the interventions, but the participant does not know how to do this.	The organization only uses qualitative research methods to measure the impact of an intervention. Qualitative research methods are preferred over quantitative research methods because they prefer measuring personal growth through qualitative methods.
<i>Participant 3</i>	The extent to which the local population is included in the peacebuilding evaluation. The extent to which the peacebuilding evaluation makes the evaluator and the local population learn something.	The attribution gap is acknowledged but attributing an intervention to (a) societal change (s) or impact(s) does not have the highest priority within peacebuilding evaluation.	A critical stance towards quantitative research methods, because ‘numbers do not say that much’ about complicated peacebuilding transformation processes and do not capture the outcomes and impacts of an intervention. Therefore, qualitative methods are required. When quantitative methods are used, it must be clarified with data gathered through qualitative methods.
<i>Participant 4</i>	The extent to which a conflict sensitive approach is integrated in the peacebuilding evaluation.	The attribution gap is perceived as ‘[...] inherent to peace building evaluations [...]’ which is an ongoing challenge without easy answers and therefore a conflict sensitive approach and the awareness of the attribution gap must be integrated in a peacebuilding evaluation.	No preference for a certain research method, but quantitative research methods are not sufficient to measure outcomes, outputs, and impacts of peacebuilding. Therefore, a mix of both methods is preferred.
<i>Participant 5 and participant 6</i>	The extent to which the evaluation shows if the realized changes in actors remain present.	The attribution gap is acknowledged, but it is ‘[...] always about contribution,’ within evaluations instead of attribution, since the peacebuilding field is so complex that it mostly is not possible to claim that changes occur only due to your own intervention(s).	There is a preference for qualitative research methods. Although quantitative methods can be helpful to track progress, it is too difficult to define or interpret the data and therefore does not add a lot of value. However, their focus is mainly on monitoring instead of evaluation, there are no clear evaluative methods defined.
<i>Participant 7</i>	The extent to which the OECD criteria and their own quality criteria are met during an evaluation.	The attribution gap is recognized. There is no clear attitude towards this gap besides that it is difficult to deal with.	No preference for a particular research method, since the method that is used, is based on the research question and the theory behind the intervention. Based on the context and the limitations of peacebuilding, it is likely that qualitative methods are used in practice, however this is not because there is a preference.

Table 5 Overview of attitudes, perspectives, and experiences of the contributing organizations

Discussion

Although the research brought about various insights into the peacebuilding evaluation field, and it seems more clear now that qualitative research methods are a necessity for effective peacebuilding evaluation; the research also suffers from limitations. This chapter does not only focus on the limitations and strengths of this thesis, it also argues how these methodological limitations could best be addressed by follow-up research. Thereafter, it suggests what academics and peacebuilding evaluators could do to expand knowledge around effective peacebuilding evaluation. This chapter also sheds some light on other insights that fell out of the scope of this thesis, due to capacity and time constraints, that could be relevant for future research.

Limitations, strengths, and methodological recommendations

Internal validity

The internal validity of the research could be increased by recruiting a larger sample. This does not rule out external factors that influence the relationship between the attribution gap and the use of qualitative research methods within evaluation, but a larger sample will be able to map more of these influences and therefore will increase knowledge on this relationship. What might have benefitted the internal validity of the research is the inclusion of organizations with different characteristics and experiences that stem from different sectors within the peacebuilding field like evaluation, humanitarian aid, development, and peacebuilding. This shows how this relationship is perceived from quite different 'angles'.

External validity

As stated in the methodology, the external validity of this research is affected by the small number of cases and the non-probability sampling methods. Both influence the ability to generalize the results of this thesis. Context-specific findings of this research, for example the criteria for effective peacebuilding evaluation relate strongly to the nature of the interventions of the organizations and can therefore not be generalized. However, the less context-specific findings of this research like mostly negative and/or reluctant attitudes towards the OECD criteria, attributing an intervention to an impact, and quantitative research methods could be generalized to the population. Another finding that could also be generalized is the necessity of qualitative research methods to attribute an intervention to an impact. This is because, as mentioned before, these organizations all have different characteristics and execute and evaluate different interventions and they still express these quite similar sentiments and preferences. This could indicate that more organizations in the population could be experiencing this.

The non-probability sampling furthermore limits the external validity since the selected cases do not have to represent the population. Future research could partly increase the external validity by trying to recruit a large sample of the population through one non-probability sampling method like convenience sampling. By increasing the sample, approaching all possible cases in a similar way, and using only selected cases based on their availability, the sample could be more representative for the population.

Reliability

Although anonymity possibly led to more consistent results and is recommended for future research, this research suffers from biases that have impacted the consistency of the data and therefore the reliability of the results. This section sheds light on how this could be avoided in follow-up research. First, the bias that is created through the subjectivity of the researcher when carrying out the discourse analysis could be addressed by making multiple people do the coding. Comparing these coded texts and choosing how to label the data based on that comparison provides a less subjective analysis and more consistent data. The experimenter bias led to inconsistent data collection. Power relations, nerves and lack of experience led to leaving some interview questions out or asking the questions in a different way than in other interviews. Examples are that not all participants were asked how their research methods

resolve the attribution gap and/or fulfil the criteria of effective peacebuilding evaluation. This makes the data collection highly inconsistent. In other words, this makes the research miss important input and present possible skewed results. This inconsistent data collection should be avoided in follow-up research. It is advised to stick to the questions of the interview protocol and ask a question a second time when a participant does not respond to the question or shows superior sentiments towards the interviewer. This would gather more useful data, make the data more comparable, and provides insight into if the participant purposefully does not answer the question or does not understand it correctly. Additionally, what also benefits the consistency of data in future research is sending the protocol to the participants beforehand. This prevents the participants to become overwhelmed by the questions. Although this does not have to be a 'bad' thing since it could show the lack of knowledge on evaluation practices, when the goal is to gather useful data, it could be advised. As an example, participants 5, 6 and 7 saw the protocol beforehand and did mention the OECD criteria quite specifically in comparison to the other participants. This shows how the data collection method could possibly have made the collected data skewed. Another reason why sending the protocol beforehand could benefit the research, is because the participants know where the interview is going which can make the collected data more comparable. It depends upon the objectives of the research, but consistency is key in providing or not providing the interview protocol before the interview. Finally, the attitudes, perspectives, and experiences of the participants are seen as representative for the organizations' points of view within this research, since most participants were appointed as most suitable by the organizations themselves to participate within the interviews. Nonetheless, it is important to keep in mind that their colleagues could have provided different data. This lowers the consistency of the data. In general, it is recommended to use multiple methods (triangulation) and include multiple participants of the organizations in follow-up research to make the data more consistent and reliable.

Recommendations for academics and organizations

Academics

To increase the knowledge around effective peacebuilding evaluation, follow-up research could take a few steps back before investigating the attribution gap since the field does not seem to be there yet. It could be beneficial to first paint a clear picture of why there is a lack of consistency in defining effective peacebuilding evaluation and why there seems to be the lack of knowledge on what evaluative methods are used by the organizations. It could also be relevant to increase understanding on why the ambitions of organizations that execute and/or evaluate peacebuilding intervention do not seem to match their actual peacebuilding evaluation practices.

The data provides way more data on the dynamics and important themes regarding peacebuilding interventions and evaluations that could be investigated but fell out the scope of this thesis. The gathered data for example showed that there are clear power relations present between the organizations. Most of the organizations collaborate through consortia and are in financial partnerships with each other. At the same time, some are each other's competition when trying to gain funding for an intervention. The power relations present between the organizations could sometimes influence the way evaluations are executed. An example is that some participants explain that they have to do evaluations in a particular way, while they do not prefer doing it in such a way. However, not executing an evaluation in a way they should, could make them unable to show accountability and hence lose funding. Power relations influence the way evaluations are executed, which could be an interesting topic to dive further into. Other themes that came up and seem to be important within peacebuilding evaluation are: (1) the popularity of including ownership and local voices within evaluation; and (2) the focus on the learning experience within an evaluation. These themes are out of the scope of this thesis but are relevant to get to know more about in future research.

Organizations

If organizations feel the need to understand and better peacebuilding evaluation, this research argues for trying to grasp where an organizations' peacebuilding evaluation currently stands and why it is like the

way it is. Does it relate to a lack of knowledge or a lack of resources? Or both? Could this be solved by introducing joint evaluations as suggested by participant 3? What are the goals of the evaluations and are they different from other organizations? Why? What is the organization's stance towards evaluation, is it a priority? Is the organization able to explain its evaluative methods into detail? Why or why not? Thereafter, to make peacebuilding evaluation more effective, it could be argued what effective peacebuilding specifically means and if the attribution gap is a priority. This then could lead to understanding what methods seem to fit this perspective best. There is still a lot of reflection that can be done by organizations that execute and/or evaluate peacebuilding interventions.

References

- Alamy Stock Photo. (2011). *Former Democratic Republic of Congo (DRC) militia members at a United Nations demobilization camp in Goma, North Kivu, before repatriation to Rwanda*. Retrieved at May 15th, 2021 at <https://bit.ly/3h2xPhg>
- Anderson, M.B., & Olson, L. (2003). *Confronting War Critical Lessons For Peace Practitioners*. Cambridge: The Collaborative for Development Action, Inc.
- Autesserre, S. (2017). International Peacebuilding and Local Success: Assumptions and Effectiveness. *International Studies Review*, 19, 114–132.
- Bamberger, M. (2012). *INTRODUCTION TO MIXED METHODS IN IMPACT EVALUATION*. Retrieved at June 21st, 2021 from <https://www.dmeforpeace.org/resource/guidance-note-3-introduction-to-mixed-methods-in-impact-evaluation/>
- Barma, N.H. (2017). Rethinking the Peacebuilding Puzzle. In N.H. Barma (Red.), *The Peacebuilding Puzzle* (pp. 11-40). Cambridge, United Kingdom: Cambridge University Press.
- Bennett, A., & Checkel, J. (2014). Process tracing. In A. Bennett & J. Checkel (Eds.), *Process Tracing: From Metaphor to Analytic Tool* (Strategies for Social Inquiry, pp. 3-38). Cambridge: Cambridge University Press.
- Blum, A. & Kawano-Chiu, M. (2012). *Learning from Nine Examples of Peacebuilding Evaluation: A Report on the 2011 Peacebuilding Evaluation Evidence Summit*. Retrieved at March 25th, 2021, from https://www.dmeforpeace.org/sites/default/files/USIP-AfP%202011%20Peacebuilding%20Evaluation%20Evidence%20Summit%20Report_FINALWEB.pdf
- Cameron, D.B., Brown, A.N., Mishra, A., Picon, M., Esper, H., Calvo, F. & Peterson, K. (2015). *Evidence for peacebuilding: an evidence gap map*. Retrieved at June 9th, 2021, from <https://www.3ieimpact.org/evidence-hub/publications/evidence-gap-maps/evidence-peacebuilding-evidence-gap-map>
- Çuhadar-Gürkaynak, E., Dayton, B., and Paffenholz, T. (2009). Evaluation in Conflict Resolution and Peacebuilding. In D.J.D. Sandole, S. Byrne, I. Sandole-Staroste, and J. Senehi (Eds.), *Handbook of Conflict Analysis and Resolution* (pp. 286-299). Oxon and New York: Routledge.
- Denskus, T. (2012). Challenging the international peacebuilding evaluation discourse with qualitative methodologies. *Evaluation and Program Planning*, 35(1), 148–153.
- Development Assistance Committee. (2007). *Encouraging Effective Evaluation Of Conflict Prevention And Peacebuilding Activities Towards DAC Guidance*. Retrieved at March 25th, 2021 from <http://www.oecd.org/development/evaluation/dcdndep/39660852.pdf>
- Gerring, J. (2017). *Case Study Research: Principles and Practices* (2nd ed., Strategies for Social Inquiry). Cambridge: Cambridge University Press.
- Hardy, C., Harley, B., & Philips, N. (2004). Discourse Analysis and Content Analysis: Two Solitudes? *Qualitative & Multi-Method Research*, 2(1), 19-22.
- Hennink, M., Hutter, I. & Bailey, A. (2011). *Qualitative Research Methods*. London: SAGE Publication.
- IISS. (2020). *Over 60% of armed conflicts ongoing for over ten years shows IISS's Armed Conflict Survey*. Retrieved at March 21st, 2021 from <https://www.iiss.org/press/2020/acs-2020>
- IISS. (2021). *IISS Armed Conflict Map*. Retrieved at June 8th, 2021, from <https://www.iiss.org/publications/armed-conflict-survey/interactive-map>
- Institute for Economics and Peace. (2018). *What is Positive Peace?* Retrieved at April 7th, 2021, from <https://positivepeace.org/what-is-positive-peace>

- Ministry of Foreign Affairs. (n.d.). *Policy and Operations Evaluation Department (IOB)*. Retrieved at June 8th, 2021, from <https://english.iob-evaluatie.nl/>
- Monnard, I. & Sriramesh, K. (2020). Public relations for peacebuilding: case study from Colombia. *Corporate Communications An International Journal*, 25(1), 48-66.
- OECD. (n.d.). *Our global reach*. Retrieved at June 8th, 2021, from <https://www.oecd.org/about/members-and-partners/>
- OECD. (2012a). Conceptual background and the need for improved approaches in situations of conflict and fragility. In OECD (Ed.), *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility Improving Learning for Results* (pp. 21-30). Paris: OECD Publishing.
- OECD. (2012b). Addressing challenges of evaluation in situations of conflict and fragility. In OECD (Ed.), *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility: Improving Learning for Results* (pp. 31-37). Paris: OECD Publishing.
- OECD. (2018). *OECD DAC Evaluation Criteria Summary of consultation responses*. Retrieved at June 8th, 2021, from https://ieg.worldbankgroup.org/sites/default/files/Data/DAC-Criteria/ConsultationReport_EvaluationCriteria.pdf
- OECD. (2019). *Better Criteria for Better Evaluation: Revised Evaluation Criteria Definitions and Principles for Use* [pdf-file]. Retrieved at June 2nd, 2021, from <https://www.oecd.org/dac/evaluation/revised-evaluation-criteria-dec-2019.pdf>
- Paffenholz, T., Abu, M., And, N. & McCandless, E. (2005). Peacebuilding and Development: Integrated Approaches to Evaluation. *Journal of Peacebuilding and Development*, 2(2), 1-5.
- Peacebuilding Evaluation Consortium (n.d.). *Guiding Steps for Peacebuilding Design, Monitoring, and Evaluation* [pdf-file]. Retrieved at June 2nd, 2021, from https://www.dmeformepeace.org/wp-content/uploads/2018/06/PEC_Guiding-Steps-for-Peacebuilding-DME.pdf
- Roser, M. (2016). *War and Peace*. Retrieved at March 21st, 2021 from <https://ourworldindata.org/war-and-peace>
- Scharbatke-Church, C. (2011). Evaluating Peacebuilding: Not Yet All It Could Be. In B. Austin, M. Fischer, H.J. Giessmann (Eds.), *Advancing Conflict Transformation The Berghof Handbook II* (pp. 459-482). Opladen/Framington Hills: Barbara Budrich Publishers.
- Schmidt, C. (2004). The Analysis of Semi-structured Interviews. In U. Flink, E. von Kardorff, & I. Steinke (Eds.), *A Companion to Qualitative Research* (pp. 253-258). London, Thousand Oaks, New Delhi: SAGE Publications.
- Smeets, S. (2018). *Content analysis and coding BS* [powerpoint-file].
- Thomas, G. (2005). The Qualitative Foundations of Political Science Methodology. *Perspectives on Politics*, 3(4), 855-866.

Appendix I: Semi-structured interview protocols

Semi-structured interview protocol English

1. What are the criteria for effective peacebuilding evaluation according to your organization?
 - a. Do these criteria differ from the criteria from ten years ago?
2. In what way does your organization experience the attribution gap?
3. What are your thoughts on quantitative research methods within peacebuilding evaluation?
 - a. How does your organization try to measure impact through quantitative research methods?
4. What are your thoughts on qualitative research methods within peacebuilding evaluation?
 - a. How does your organization try to measure impact through qualitative research methods?
5. What research methods are preferred by your organization within peacebuilding evaluation to capture the impact of the peacebuilding practices?
 - a. How are these research methods expected to help resolve the attribution problem/gap?
6. To what extent do these methods in practice fulfil the criteria of effective peacebuilding evaluation mentioned above?
7. Are there other insights or obstacles regarding this topic that you find relevant to share?

Semi-structured interview protocol Dutch

1. Wat zijn de criteria voor effectieve peacebuilding evaluatie volgens jouw organisatie?
 - a. Verschillen deze criteria van de criteria die tien jaar geleden werden gehanteerd?
2. Op wat voor manier ervaart jouw organisatie de attribution gap?
3. Hoe kijk jij naar kwantitatieve onderzoeksmethoden binnen peacebuilding evaluaties?
 - a. Hoe probeert jouw organisatie impact te meten middels kwantitatieve onderzoeksmethoden?
4. Hoe kijk jij naar kwalitatieve onderzoeksmethoden binnen peacebuilding evaluaties?
 - a. Hoe probeert jouw organisatie impact te meten middels kwalitatieve onderzoeksmethoden?
5. Welke onderzoeksmethoden hebben de voorkeur binnen jouw organisatie in peacebuilding evaluatie om impact van de peacebuilding projecten te meten?
 - a. Hoe wordt verwacht dat deze onderzoeksmethoden de attribution gap oplossen?
6. In welke mate vervullen deze methoden de eerdergenoemde criteria van effectieve peacebuilding evaluatie?
7. Zijn er nog andere dingen of obstakels gerelateerd aan dit onderwerp die je relevant vindt om te noemen?

Appendix II: Coding scheme

1. Criteria effective peacebuilding evaluation

Variables	Indicators
<i>Relevance</i>	The extent to which the intervention objectives and design respond to beneficiaries', global, country, and partner/institution needs, policies, and priorities, and continue to do so if circumstances change.
<i>Coherence</i>	The compatibility of the intervention with other interventions in a country, sector, or institution.
<i>Effectiveness</i>	The extent to which the intervention achieved, or is expected to achieve, its objectives, and its results, including any differential results across groups.
<i>Efficiency</i>	The extent to which the intervention delivers, or is likely to deliver, results in an economic and timely way.
<i>Impact</i>	The extent to which the intervention has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects.
<i>Sustainability</i>	The extent to which the net benefits of the intervention continue or are likely to continue.
<i>Other</i>	The extent to which other criteria are formulated for effective peacebuilding evaluation.

(OECD, 2019, pp. 7-12)

2. Attribution gap within peacebuilding evaluation

Variables	Indicators
<i>Experience with (the difficulty of) attributing a peacebuilding intervention to (a) societal change(s) or impact(s)</i>	The way (the difficulty of) attributing a peacebuilding intervention to (a) societal change(s) or impact(s) is experienced.
<i>Attitude towards (the difficulty of) attributing a peacebuilding intervention to (a) societal change(s) or impact(s)</i>	The way (the difficulty of) attributing a peacebuilding intervention to (a) societal change(s) or impact(s) is viewed.

3. Research methods within peacebuilding evaluation

Variables	Indicators
<i>Attitude towards quantitative research methods</i>	<ul style="list-style-type: none"> The way quantitative research methods are viewed. The extent to which quantitative research methods are preferred over qualitative research methods. The ways in which quantitative research methods are argued to resolve the attribution gap. The ways in which quantitative research methods are argued to fulfil the criteria of effective peacebuilding evaluation.
<i>Use of quantitative research</i>	<ul style="list-style-type: none"> The extent to which the goal of the research method is to 'measure, count or quantify,' (Hennink et al., 2011, pp. 16-17).

<i>methods to measure impact</i>	<ul style="list-style-type: none"> The extent to which data is gathered through ‘population surveys, opinion polls, exit interviews,’ (Hennink et al., 2011, pp. 16-17).
<i>Attitude towards qualitative research methods</i>	<ul style="list-style-type: none"> The way qualitative research methods are viewed. The extent to which qualitative research methods are preferred over quantitative research methods. The ways in which qualitative research methods are argued to resolve the attribution gap. The ways in which qualitative research methods are argued to fulfil the criteria of effective peacebuilding evaluation.
<i>Use of qualitative research methods to measure impact</i>	<ul style="list-style-type: none"> The extent to which the goal of the research method is to understand the context, the process and/or the influences (Hennink et al., 2011, pp. 16-17). The extent to which data is gathered through ‘in-depth interviews, observation, group discussions,’ (Hennink et al., 2011, pp. 16-17).

4. Other

Starting Variable	Eventual variables	Indicator
<i>Other input regarding to effective peacebuilding evaluation</i>	<i>Localization in peacebuilding evaluation</i>	The extent to which the local population is included in peacebuilding evaluation.
	<i>Emphasis on learning within peacebuilding evaluation</i>	The extent to which there is an emphasis on learning within peacebuilding evaluation.
	<i>Mentioning of other actors regarding peacebuilding evaluation</i>	The extent to which other actors are mentioned regarding peacebuilding evaluation.
	<i>Resource difficulties regarding peacebuilding evaluation</i>	The extent to which peacebuilding evaluations are linked to resource difficulties.
	<i>Contribution instead of attribution within peacebuilding evaluation</i>	The extent to which the focus is on contribution instead of attribution of a peacebuilding intervention.