Negotiating effectiveness: The politics of results in advocacy for development and its evaluation

Bodille Arensman

Negotiating effectiveness:

The politics of results in advocacy for development and its evaluation

Bodille Arensman

Thesis committee

Promotor

Prof. Dr D.J.M. Hilhorst

Professor of Humanitarian Aid and Reconstruction

International Institute of Social Studies Erasmus University

Co-promotor

Dr M.G.J. van Wessel

Assistant professor, Strategic Communication Group

Wageningen University & Research

Other members

Prof. Dr C. Leeuwis, Wageningen University & Research

Dr I. Guyt, Oxfam Great Britain, Oxford, UK

Dr H. Huyse, KU Leuven, Belgium

Dr P.G.H. Engel, School of Management Maastricht

This research was conducted under the auspices of the Wageningen School of Social Sciences (WASS)

Negotiating effectiveness:

The politics of results in advocacy for development and its evaluation

Bodille Arensman

Thesis

submitted in fulfilment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus

Prof. Dr A. P. J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Monday 18 June 2018

at 1.30 p.m. in the Aula.

Bodille Arensman

Negotiating effectiveness:

The politics of results in advocacy for development and its evaluation,

206 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2018)

With references, with summary in English

ISBN: 978-94-6343-872-8

DOI: 10.18174/448390

To my loving family who believe in me and all my adventures.

Acknowledgements

I have long aspired to do research and 'do my PhD'. This started during my university studies and continued during my first year working in the 'real' world. However, the chance to finally start my PhD came as one of these seize the moment opportunities to which I am very grateful. In July 2012, I was ecstatic to start my research as part of an evaluation consortium. I could now not only start devoting my time to elaborately explore the effectiveness of advocacy in international development organisations, but also contribute to the practices of advocacy evaluation. This resulted in an amazing journey with ups and downs, including steep learning curves, lots of travels, physically as well as intellectually. An incredibly inspiring journey, most of all because of to the people that guided me on my way and that made me part of their personal stories.

First and foremost, this study would never have been possible without the support of many staff members of the various development organisations and networks that were involved in the MFS II ILA Evaluation. I would therefore like to thank the staff of these organisations and their partners worldwide. Without the openness and support of the staff and their partners in these organisations, I would never have been able to gather the insights that are included in this doctoral dissertation. The organisations with whom I worked most intense with were the African Child Policy Forum (ACPF), the Global Partnership for the Prevention of Armed Conflict (GPPAC) and the Dutch development organisation Cordaid. I am grateful for all the moments of mutual learning over the past years, and I would like to thank each of the staff members of these organisations for their kind openness to have me observe their day-to-day processes in the midst of their advocacy engagements. I would also like to thank the MFS II Evaluation Team for the opportunity of having a PhD included in the evaluation processes. With a special thanks to Jennifer Barrett, with whom I had many important discussions during the evaluation. Jennifer provided critical input on my analyses, always challenging me to clarify my meaning. She also edited the different chapters.

To my supervisors, Thea Hilhorst and Margit van Wessel, I extend my deep gratitude for believing in my abilities, for inspiring and motivating me to continue my research and go on with my work when sometimes it was challenging. Most of all, they continued to critically question my findings and arguments, forcing me to be more clear and more to the point. Thea, your warm openness, the diversity of discussion topics and your often unconventional ways taught me the joy of doing research and inspired me to think more outside of the box. Margit, your door was always open and you tirelessly read and discussed my drafts and ideas. With

your critical questions you always triggered me to dig deeper, to be more bold, and to challenge myself. Meanwhile, your sense of humour always kept me laughing as well. Both of you are a great inspiration to me, and under your supervision I learned a lot and I developed into a critical academic and evaluator.

Thanks also to my colleagues with whom I had many interesting discussions, Cornélie, José, Carolien, Bart, and Wendy. Thank you for all your support during the past years. A special thanks to José with whom I mostly had WhatsApp contact bridging the distance between DR Congo and Rotterdam. We send a lot of texts blowing off steam on the tough PhD process, sharing our concerns and triumphs while inspiring each other to continue and keep working on our analyses and chapters. I am glad I have come to know you as a colleague and a friend. Also, a very special thanks to Cornélie, my evaluation and PhD writing buddy. We grew close during our travels to Ethiopia, and we continued seeing each other regularly afterwards during weekly writing sessions. I never grow tired of our discussions and I have come to know you as an inspirational person. You taught me a great deal about life, and inspired me to be intrigued. Over lunch during our weekly sessions, we vividly discuss the most diverse topics often digging deep into the critical issues of the international development world. When I felt stuck you were always there for me to inspire me to continue, and I only hope that I do the same for you. I have come to know you more as a friend than a colleague, and I am looking forward to continue our discussions and friendship also after we both finished our PhDs.

Coming closer to home, I am very grateful to the people that were there for me long before I started my PhD. Who contributed to the most important foundations that eventually led to finalizing this dissertation. I am grateful to my dear family and friends. While most of you have only a vague idea of what on earth I have been doing for the past years, spending all those nights and weekends working and cancelling on appointments or sometimes disappearing for weeks and months on end, you never stopped believing in me. Be it with supporting words of wisdom and strength, listening to my worries and concerns, or by having great nights out or just by letting me be in my PhD bubble. The most important and my anchor in life is my family. I am grateful to my parents for all the opportunities they gave me together with their tireless support and their belief in my abilities as a person, and as an academic. From my parents I inherited my strong sense of perseverance and justice. I would never have come to this point in my life if it weren't for them.

My wonderful and creative mother, Els. You have always been there for me. You supported me no matter what, mostly by giving me loads of love. You are always eager to know what I do or what I am going through. Thank you for providing a listening ear and a curious

mind. You translated my findings in a painting that is now the cover of this dissertation. I am so proud of who you are. My dear and loving father, Leo. You are always there for me with advise, and with a hug. Over the years, you did not always understand why I was doing what I did, but I always felt you were proud of me. You gave me every opportunity to blossom into the person I am today. I could not have wished for better parents, you both taught me to be independent, loving, strong and honest, with a hint of stubborn and always with a good sense of humour. Thank you both for being my safe haven. To my wonderful sister Liselot, you have supported me greatly over the years and you never doubted my abilities for a second, albeit you have been worried about my intense work ethic. You supported me by letting me be myself, by surprising me with fun evenings with great food, and by insisting that we take time off for relaxing yoga weekends. Somehow, we always end up critically analysing our lives and psyche. You are always there for me, sometimes critically questioning me, other times by overloading me with fun as we have our laughter attacks. Thank you for being there for me. And finally, to my dear grandma, who is always asking me about my travels and my findings while feeding me home-made cookies and coffee. You have been very important in my life, and with your 94 years you are still an inspiration to me the way you are: the way you always know everything often with a wicked sense of humour. To my partner, Niels, I have come to know you during the final stretch of my PhD, which is known to be the hardest. You have stood by me and supported me in numerous wonderful ways that I can't begin to describe, and that I will for ever be grateful for. Thank you all for being there for me in so many wonderful ways. I am looking forward to what other adventures life has in store for me, and I gladly dedicate this thesis to you, my dear family.

Table of Contents

Acknowledgements **6**
List of Tables and Figures **12**
List of abbreviations **13**

1. Introduction: The Puzzle of Advocacy for Development and its Evaluation **15**
1.1. Introduction 15
1.2. The MFS II ILA evaluation 19
1.3. Advocacy and its evaluation: A challenging puzzle 23
1.3.1. Conceptualising advocacy 23
1.3.2. Challenges in evaluating advocacy 26
1.3.3. Conceptualising advocacy effectiveness 27
1.4. Research questions and objective 30
1.5. Theoretical and analytical framework 31
1.5.1. Conceptualising the politics of results 31
1.5.2. Results-based evaluation traditions in international development 33
1.5.3. Meaning-making in advocacy evaluation and the politics of results 35
1.6. Methodology 37
1.6.1. An iterative approach 38
1.6.2. Methodological design and data gathering 39
1.6.3. Wearing different hats 42
1.7. Outline of this dissertation 43
2. Does local ownership bring about effectiveness? The case of a transnational advocacy network **47**
2.1. Introduction 48
2.2. Effectiveness and ownership in development 49
2.3. Effectiveness and ownership in advocacy networks 51
2.4. Methods 53
2.4.1. Case Study 54
2.5. Findings 55
2.5.1. 'Being an inclusive network' 56
2.5.2. Accountability as an inclusive network 57
2.5.3. The inclusive advocacy message 59

2.6. Being inclusive and effective ... 61
2.7. Conclusions ... 64
3. Negotiating effectiveness in transnational advocacy evaluation ... **67**
3.1. Introduction ... 68
3.2. Advocacy challenges realist evaluation ... 69
3.2.1. The results-oriented approach ... 69
3.2.2. Advocacy evaluation as socially negotiated ... 71
3.3. Methods and case ... 72
3.3.1. Creating space for interaction ... 73
3.4. Shaping the evaluation process and outcomes ... 75
3.4.1. Identifying outcomes ... 75
3.4.2. Measuring outcomes ... 78
3.4.3. Presenting outcomes ... 82
3.5. Discussion ... 87
3.6. Conclusion ... 88
4. Twinning 'Practices of Change' with 'Theory of Change': Room for Emergence in Advocacy Evaluation ... **91**
4.1. Introduction ... 92
4.2. Theory of Change and Advocacy Evaluation ... 94
4.3. Creating Space for 'Practices of Change' ... 96
4.4. Methods and Case Study ... 99
4.4.1. The Case Study ... 100
4.5. Findings ... 101
4.5.1. Theory and 'Practices of Change' ... 101
4.5.2. Intended and Emergent Outcomes ... 104
4.5.3. Recursive Outcome Loops ... 105
4.6. 'Practices of Change' ... 107
4.7. Conclusion ... 110
5. Advocacy outcomes are not self-evident: the quest for outcome identification ... **113**
5.1. Introduction ... 114
5.2. Evaluating advocacy outcomes ... 116
5.3. Limitations of the existing methods ... 119
5.4. Case studies, data collection and analysis ... 120
5.4.1. Advocacy for development evaluation. ... 120
5.4.2. Data collection and analysis ... 121
5.5. The quest for advocacy outcomes ... 122

5.5.1. Room for interpretation 122
5.5.2. Outcomes are constructed 127
5.5.3. The politics of outcomes 130
5.6. Discussion 134
5.7. Conclusions 137
6. Putting the puzzle together: Conclusions and Contributions 139
6.1. Introduction to the conclusion 139
6.2. Main findings: Conclusions from the four empirical chapters 140
6.2.1. Advocacy effectiveness is multidimensional 142
6.2.2. Advocacy effectiveness is negotiated 144
6.2.3. Theory of change needs to be twinned with practices of change 145
6.2.4. Advocacy effectiveness is not self-evident 147
6.3. Main findings: Cross-cutting conclusions 149
6.3.1. Formal and informal meaning-making 149
6.3.2. The politics of results is multidimensional 151
6.4. Main findings: Practical relevance and directions for future research 152
6.4.1. Advocacy evaluation demands a different approach 152
6.4.2. A shift in the international development world 155
6.4.3. Steps forward: A future research and practice agenda 156
Summary 159
References 164
Annex 1 177
Annex 2 197
Annex 3 198

List of Tables and Figures

Figure 1 - What is advocacy? 25

Figure 2 - The diverse dimensions in understanding advocacy effectiveness 26

Figure 3 - ACPF theory of change 103

Figure 4 - ACPF practice of change 104

Table 1 - Overview of the advocacy alliances and programmes in the Dutch Co-Financing Evaluation of International Lobby and Advocacy (MFS II ILA evaluation) 22

Table 2 - Third feedback round of the advocacy evaluation, January 2015 29

Table 3 - Third feedback round of the advocacy evaluation, January 2015 82

List of abbreviations

ACPF	African Child Policy Forum, Addis Ababa
AU	African Union
CSO	Civil society organisation
CSW	Commission on the Status of Women
CoC	Communities of change
DFID	Department for International Development (UK)
EA	Ecosystem Alliance
EU	European Union
FGG	Fair Green and Global Alliance
FFF	Freedom from Fear Alliance
GPPAC	Global Partnership for the Prevention of Armed Conflict
ICAI	Independent Commission for Aid Impact
ILA	International lobbying and advocacy
INGO	International non-governmental organisation
ISG	International steering group
M&E	Monitoring and evaluation
MDGs	Millennium Development Goals
MFS	Netherlands Co-financing System (*Medefinancieringsstelsel*)
MFS II	Netherlands Co-financing System (*Medefinancieringsstelsel*), 2nd term (2011–2015)
MFA	Ministry of Foreign Affairs
NAP	National Action Plan
NAP 1325	National Action Plan for Women, Peace and Security
NATO	North Atlantic Treaty Organisation
NGO	Non-governmental organisation
NL	The Netherlands
NL MFA	Netherlands Ministry of Foreign Affairs
NWO	Netherlands Organisation for Scientific Research
ODI	Overseas Development Institute (UK)
OECD (DAC)	Organisation for Economic Co-operation and Development (Development Assistence Committee)
OH	Outcome harvesting
OM	Outcome mapping
PRAs	Priority result areas
RIGO	Regional intergovernmental organisation
SAP	Strategy-as-practice
SGE	Stichting Gezamenlijke Evaluaties/Foundation for Joint Evaluations
SRHR	Sexual and reproductive health and rights
TAN	Transnational Advocacy Network
T4C	Together 4 Change Alliance
ToC	Theory of change
ToR	Terms of reference
TTIP	Transatlantic Trade and Investment Partnership
UN	United Nations
UNCRC	United National Committee on the Rights of the Child

UNSCR	United Nations Security Council Resolution
UNSCR 1325	United Nations Security Council Resolution on Women, Peace and Security
UNSG	United Nations Secretary General
UN WOMEN	United Nations Entity for Gender Equality and the Empowerment of Women
US(A)	United States (of America)
USAID	United States Agency for International Development
Wageningen UR	Wageningen University & Research
WLPS	Women's Leadership for Peace and Security Program (Cordaid)
WG	Working Group

1. Introduction: The Puzzle of Advocacy for Development and its Evaluation

1.1. Introduction

Increasingly, non-governmental organisations (NGOs) engaged in development consider advocacy an important strategy for structural development change.[1] Advocacy for development is a growing trend, especially with the rising numbers of development organisations considering the limitations to the effect of service delivery in international development on the achievement of structural change. Key players increasingly realise that the pursuit and achievement of structural change and development are political processes that require strategies such as advocacy (Hudson, 2002; Jolly et al., 2004; Jones, 2011; Keck and Sikkink, 1998; Kremer et al., 2009; Teles and Schmitt, 2011).

This trend towards advocacy for development is occurring at a time when organisations in the public domain, such as NGOs, need to prove their worth by showing how effective they are. In other words, they need to prove their effectiveness by accounting for funding spent in terms of results achieved. NGOs and donors have been under scrutiny for years, facing criticism for the millions of euros in funding spent that, according to critical scholars, did not always lead to specific, immediate and clear results that could be used to prove the organisations' effectiveness (Best, 2017; Easterly, 2006; Kremer et al., 2009). Organisations are considered effective when they make visible the results they predefined and achieved. 'Was the funding spent effectively?'

[1] NGOs, but also donors engaged in development increasingly focus on advocacy as strategy for change, see Hudson, 2002; Morariu and Brennan, 2009; Netherlands Ministry of Foreign Affairs, 2015; Raynor et al., 2009; Tsui et al., 2014; UNICEF, 2010. In 2015, the Netherlands Ministry of Foreign Affairs (NL MFA) made advocacy the main strategic policy focus of the development aid by the NL MFA, with a maximum of 925 million euros earmarked for NGO advocacy over a period of five years. With its co-financing system ('Dialogue and dissent: Strategic partnerships for lobby and advocacy'), the NL MFA financially supported only those programmes that included advocacy as one of their main strategies for change. See 'Ploumen vernieuwt samenwerking met ontwikkelingsorganisaties', a news item on the NL MFA's website at https://www.rijksoverheid.nl/actueel/nieuws/2015/01/30/ploumen-vernieuwt-samenwerking-met-ontwikkelingsorganisaties (30 Jan 2015). See also the policy framework 'Dialogue and dissent: Strategic partnerships for lobbying and advocacy' at https://www.government.nl/documents/regulations/2014/05/13/policy-framework-dialogue-and-dissent (NL MFA, AVT14/BZ111606A, 13 May 2014). Also the Swedish Development Agency SIDA works on active advocacy for development change and the implementation of the Swedish Government's development policy, stating: "Dialogue and advocacy are important tools throughout development cooperation and synergies and complementarity are sought."(SIDA, 2016: 46). See the policy framework for Swedish development cooperation and humanitarian assistance 2016/17:60: http://www.government.se/49a184/contentassets/43972c7f81c34d51a82e6a7502860895/skr-60-engelsk-version_web.pdf

has become a central question (Kremer et al., 2009), demanding measurable, visible and tangible results (Eyben, 2015; Riddell, 2014). However, advocacy for development does not easily lend itself to answering effectiveness questions in such measurable terms. Despite the growing practice of NGO advocacy and an equally increasing literature on this advocacy its evaluation, establishing effectiveness remains problematic.

Although the literature provides important insights, it does not touch upon how advocacy effectiveness is given meaning or how this is affected by the politics involved around results, regarding the accountability, credibility and legitimacy of the organisation. There is a gap in the literature pertaining to the specific challenges of the effectiveness of advocacy for development and its measurement. With advocacy gaining ever more ground in the development world, these issues are becoming increasingly relevant. This doctoral dissertation therefore asks the following key question: *How does the practice of advocacy for development give meaning to effectiveness, and how do the politics of results play out in the evaluation of advocacy?* Here, effectiveness is understood as the processes and practices through which an intervention reaches its goals. Hence, it is not about asking *whether* an intervention reached its goals, but rather *how* these goals were reached by an intervention. The determination of whether a course of action was effective is not self-evident, and this dissertation understands the meaning of 'effectiveness' to be determined through a process of human interaction that is intrinsically political.

Although the effectiveness of development is notoriously difficult to measure, this is even more difficult with the often intangible outcomes of advocacy. How can advocacy be effective, and how can this be measured? There is an assumption among NGOs and donors that advocacy will result in structural changes, thus increasing development effectiveness (Hudson, 2002; Morariu and Brennan, 2009; Netherlands Ministry of Foreign Affairs, 2015; Raynor et al., 2009; Tsui et al., 2014; UNICEF, 2010). However, actually proving effectiveness is especially difficult in the context of advocacy for development, partially because structural changes in international development often take a very long time to develop and involve many diverse stakeholders operating at multiple levels and across borders[2] (Arensman et al., 2015; Barrett et al., 2016). Embedded in the quest to be more effective is the development of models that help to predefine and predict effectiveness, pressuring organisations to show measurable results within funding periods that range from a couple of months to approximately five years (Eyben

[2] For example, women's rights and equal marriage rights for gay and lesbian couples took many years of diverse activists, social movements, organisations, governments and civil society actors pursuing changes in policies, practices, behaviour and awareness of the issues.

et al., 2015; Riddell, 2014). This influences the development world because the tendency for measurable effectiveness leaves its mark on organisational procedures, protocols and interests, affecting relations, politics and interactions (Best, 2017; Eyben, 2015; Guljarani, 2011).

Because it is difficult to establish the outcomes of advocacy (and even more so the effectiveness of efforts to reach these outcomes), the process of defining outcomes and effectiveness is profoundly political. I refer to this political process as the *politics of results*—the negotiated process by which outcomes and effectiveness are given meaning in the practice of advocacy for development. Because the outcomes and effectiveness of advocacy for development are so difficult to operationalise, in practice, it turns out that the politics of results often boils down to the question of whether the intervention or organisation has succeeded in achieving legitimacy and credibility in the eyes of other stakeholders.

Advocacy presents challenges for the evaluation of effectiveness, especially because of advocacy's dynamic character and because of the substantive meaning of effectiveness in the contexts where advocates and advocacy evaluators operate (i.e. *politics of results*). Advocacy for development is mostly about abstract ideals that are not necessarily achieved within funding cycles, or maybe even ever (e.g. a world free of human rights violations or conflict). Advocates often seek change on issues that are intertwined with moral principles (e.g. basic human rights, the right to clean water, the right to justice, distribution of land, protection from violence, questioning power structures, fair trade). When calling for support, advocates appeal to shared moral standpoints, making propositions with regard to the worthiness of their efforts or the significance of their success. In pursuing structural changes, advocates need to navigate, manoeuvre and strategize in changing circumstances. The results of these actions, however, are not necessarily obvious or even visible. How effectiveness is given meaning in the context of such processes is thus challenging for evaluators.

Even though advocacy effectiveness and its evaluation are increasingly taken up by researchers, the discussion of advocacy effectiveness remains rather limited. Numerous studies have focused on explaining advocacy effectiveness by studying the success and failure of policy changes, considering the importance of a resonating message (Keck and Sikkink, 1998), the political salience of the advocacy message and campaign (Wong, 2012), and the network density and strength in balancing the collective and the individual (Keck and Sikkink, 1998; Saz-Caranza, 2012). Brown, Ebrahim and Batliwala (2011: 1104), in a comparative analysis of 10 advocacy NGOs, defined advocacy effectiveness in terms of policy influence, organisational capacity building, strengthened coalitions and networks, and widened public awareness. They found that the structures in which advocacy NGOs work (networks, coalitions and partnerships)

enable or constrain decision making and strategizing, leading to success or failure. Other studies exploring the achievement of effectiveness in advocacy have pointed out that planning, monitoring and evaluation are essential and have emphasised the importance of advocates using a theory of change (ToC) to understand the contexts in which they move and learn (Beer and Reed, 2009; Coffman, 2007; Gardner and Brindis, 2017; James, 2011). In these studies, effectiveness is given meaning in relation to organisational capacity, structure, size and strength (Alliance for Justice, 2015; Shawki, 2011), as well as the institutional support for and organisational commitment to advocacy (Skene-Pratt, 2013; Wong, 2012).

The existing literature does not discuss what constitutes the meaning of effectiveness (*Whose effectiveness? What effectiveness? How does effectiveness relate to the politics involved?*) in the contexts in which advocates and advocacy evaluators operate. Effectiveness is often understood as the achievement of predefined goals. According to the studies discussed above, elements such as organisational strength and size, issue resonance, and network density play an important role in explaining the successes and failures of advocacy campaigns. However, in advocacy for development, the same elements can have different results (see Keck and Sikkink, 1998; Shawki, 2010, 2011). Success or failure is not always easily traced, observed or measured, and results cannot always be predicted or predefined. The meaning of effectiveness comprises more than just the successful achievement of intended goals and needs to be understood as inherently part of everyday practices and politics. Therefore, meaning-making is a central element in this doctoral dissertation.

With this research, I will answer the overarching questions of how the practice of advocacy for development gives meaning to effectiveness and how the politics of results plays out in the evaluation of advocacy. As a researcher and evaluator, I was involved in a transnational advocacy evaluation from 2012 to 2015. Evaluating advocacy for development, I experienced first-hand the challenges of effectiveness in transnational advocacy and its measurement, and I witnessed the role played by the meaning given to effectiveness in the politics of results. The evaluation was part of the second Dutch Co-Financing System (MFS II) 2010–2015 Evaluation of International Lobbying and Advocacy, abbreviated as the MFS II ILA evaluation. The case studies for the research presented in this dissertation cover evaluated advocacy for development programmes as well as the process of evaluating these programmes.

This dissertation consists of three components: an introduction, four chapters presenting the results, and a conclusion summarising the findings and contributions. The present chapter begins with an introduction to the case studies used in this dissertation, explaining the scope and aims of the evaluation. I then discuss the discourses that are important for understanding

advocacy for development and the challenges to evaluating its effectiveness, including the limitations of the evaluation practices that I encountered. This contextualising account is followed by an explanation of the research questions. Finally, I draw out the theoretical and analytical framework that sets the frame of reference for understanding the politics of results and the role of meaning-making, followed by an account of the methodology.[3] The second part of the dissertation consists of the four empirical chapters that present my findings.[4] These four chapters unpack how the practices of advocacy for development give meaning to effectiveness in advocacy practices and discuss how the politics of results plays out in the evaluation of advocacy in the quest to pursue, negotiate, understand and interpret effectiveness. Each of these chapters addresses different audiences and engages in different discussions based on diverse literature drawn from development studies, evaluation studies, advocacy studies and organisation studies. The third section is the conclusion, which discusses the overarching findings and contributions of my research.

1.2. The MFS II ILA evaluation

For this doctoral dissertation I studied advocacy for development and its evaluation as part of one of the largest transnational advocacy evaluation programmes worldwide. The evaluation measured two moments in time (2012 and 2014), creating a multiple-year programme that enabled the evaluating team—with permission from the donor and stakeholders—to weave a PhD project into the process. I had the dual role of both an evaluator and a PhD researcher.

This evaluation of international lobbying and advocacy programmes was financed by the second Dutch Co-Financing System (*Medefinancieringsstelsel*, or 'MFS II'), which was the Netherlands Ministry of Foreign Affairs' (NL MFA) 2011–2015 grant framework for co-financing agencies.[5] Under MFS II, the NL MFA awarded a total of 1.9 billion euros to 20 alliances of Dutch development organisations. The MFS II framework stipulated that each alliance carry out independent external evaluations to be able to make valid evaluative statements about the effective use of the available funding. To meet these evaluation

[3] The empirical chapters also provide an account of the methodology. However, this is discussed more thoroughly in this introductory chapter because the word limits imposed by most academic journals prevented the inclusion of elaborate accounts in the separate chapters.
[4] These are four articles, including three that are published and one that has been submitted and is currently under review. The three published articles appear in *Third World Quarterly*, *Evaluation* and the *American Journal of Evaluation*.
[5] For a summary of the evaluation process and findings, see Annex 1.

requirements, a joint evaluation programme was developed and approved by the NL MFA, with the overall purpose for evaluating MFS II-funded development interventions to account for the results and to contribute to the improvement of future development interventions. The alliances formed the Foundation for Joint Evaluations (*Stichting Gemeenschappelijke Evaluaties*, SGE) to act on behalf of the evaluated organisations. The SGE was assisted by an internal reference group, consisting of seven evaluation experts. The joint evaluation was administered by the Netherlands Organisation for Scientific Research (NWO) on behalf of the NL MFA.

MFS II succeeded MFS I (2005–2010) that focused on strengthening civil society in the global South to achieve structural poverty alleviation.[6] MFS II was characterised by its three-fold strategy under the overarching objective of structural poverty alleviation: achieving the Millennium Development Goals (MDGs), strengthening civil society and capacity building, and setting the international agenda and changing policy and practice.[7] To assess the effectiveness and efficiency of these strategies, three types of evaluation studies were conducted:

1. Evaluations at country level, looking at the capacity strengthening of Southern partner organisations, the achievement of the MDGs and efforts to strengthen civil society (with a total budget of 9,298,157 euros);
2. A synthesis study of these country-level evaluations (with a budget of 675,000 euros); and
3. An evaluation of international lobbying and advocacy (with a budget of 810,000 euros).[8]

The evaluations had three specific aims. The first aim was to analyse and assess the effectiveness, efficiency and relevance of the development interventions per strategy, asking what was achieved in terms of changes and impact. The second aim was to provide justified recommendations to enable Dutch development organisations and the NL MFA to draw lessons for future interventions. The third aim was to contribute to methodology development. In the country-level evaluations, the methodological aim was to compare different methodologies of

[6] Policy document: A.M.A. van Ardenne-van der Hoeven (2004) Brief van de minister voor ontwikkelingssamenwerking. 22 December (27 433, nr. 29): https://zoek.officielebekendmakingen.nl/kst-27433-29.html (15-03-2017).
[7] The guiding policy document is entitled 'Maatschappelijke organisaties: Samenwerken, Maatwerk, Meerwaarde' (31 933, nr. 1), discussed in the Netherlands Parliament 'Algemeen Overleg' 20 May 2009 (31 933, nr. 3).
[8] NWO. 'What is MFS II?' http://www.nwo.nl/en/research-and-results/programmes/Joint+MFS+II+Evaluations+of+development+interventions/background (15-03-2017).

evaluation. In the MFS II ILA evaluation, this aim meant developing and applying new and innovative methodologies for evaluating advocacy. The MFS II evaluations were thus set up with a broad contextual scope in terms of programmes, focus areas, and evaluation aims and terms.

This doctoral dissertation specifically focuses on the MFS II ILA evaluation, studying advocacy for development and its evaluation. The MFS II ILA evaluation covered eight transnational advocacy programmes, implemented by a variety of alliances of Dutch development organisations. These eight advocacy programmes were clustered in three thematic focus areas (see Table 1): sustainable livelihoods and economic justice (four programmes); sexual and reproductive health and rights (one programme); and protection, human security and conflict prevention (three programmes). Although they were clustered under these broad themes, the programmes were diverse in focus, scope and organisational character. The programmes included global and regional networks; Dutch, international and Southern development organisations; and policy, advocacy and research organisations. Programmes included campaigns, networks, activism, lobby and partnerships. The covered themes were also wide-ranging, including international development agenda issues such as the Sustainable Development Goals, the New Deal, United Nations Security Council Resolution 1325 on gender equality, and the United Nations Convention on Climate Change. Other programmes focused on trade-related issues such as the Transatlantic Trade and Investment Partnership. There were also programmes focusing on issues at the national level, such as Dutch policies and financing for sexual and reproductive health and rights, ecosystem commodities in Asia (biofuels, palm oil and soy), and local women's networks. These examples serve to illustrate the diversity in terms of advocacy focus and organisational structure.

For the MFS II ILA evaluation, three priority result areas of international lobbying and advocacy were defined by the NL MFA: *agenda setting, policy influencing and changing practice*. These priority result areas were defined according to phases that most lobbying and advocacy programmes presumably go through to reach their objectives and results (MFS II ILA, Terms of Reference, 2012). However, it was stressed that these phases are not linear and are often iterative and complex.

The programmes were to be evaluated by answering five main evaluation questions:

1. What are the changes achieved in the three priority result areas (agenda setting, policy influencing, changing practice) through international lobbying and advocacy on the thematic clusters (…) during the 2012–2014 period?

2. Do the international lobbying and advocacy efforts of the MFS II alliances and their partners contribute to the identified changes (i.e. measuring effectiveness)?
3. What is the relevance of these changes?
4. Were the efforts of the MFS II alliances efficient?
5. What factors explain the findings drawn from the questions above?

The evaluation focused on being results-oriented, learning-oriented and analysis-oriented (see also Annex I for a summary of the evaluation findings and lessons learned).

Table 1. Overview of the advocacy alliances and programmes in the Dutch Co-Financing Evaluation of International Lobby and Advocacy (MFS II ILA evaluation)

Thematic area	*Alliance*	*Aims of the evaluated advocacy programmes*
Sustainable livelihoods and economic justice	Ecosystem Alliance (EA)	EA aimed to strengthen livelihoods and ecosystems in developing countries (palm oil, biofuels, soy).
	Fair, Green and Global Alliance (FGG)	FGG aimed to contribute to reducing poverty and encouraging socially just and environmentally sustainable development.
	Hivos Alliance, People Unlimited 4.1	Hivos carried out an international lobby and advocacy programme that aimed for 100% sustainable energy as quickly as possible at global level.
	Impact Alliance (IA)	IA was implemented as part of the GROW Campaign, aimed at creating a better future where everyone has enough to eat.
Sexual and reproductive health and rights	SRHR Alliance	The SRHR Alliance worked towards equal access to sexual and reproductive health information and services, as well as all people having the freedom to make informed choices about their sexuality.
Protection, human security and conflict prevention	Communities of Change (CoC)	CoC advocacy focused on increasing women's inclusion, women's participation and gender equality in peace and security processes, policies and practices.
	Freedom from Fear (FFF)	FFF focused on human security and human rights, networking for conflict prevention and peacebuilding aimed to influence and mobilise the United Nations, regional intergovernmental organisations and state actors, moving away from reaction towards prevention of conflict.
	Together4Change	The evaluation focused on the Ethiopia-based African Child Policy Forum (ACPF), a Southern partner in the Alliance that executed the international lobbying and advocacy component of the programme in Africa. ACPF advocated for pan-African child rights.

1.3. Advocacy and its evaluation: A challenging puzzle

Establishing advocacy effectiveness challenges the narrow focus of results-driven evaluation methods. To explain this, I will first discuss advocacy, before exploring the question of why advocacy is difficult to evaluate. This will position the discussion of advocacy effectiveness in the broader context and show how the existing literature defines advocacy effectiveness.

1.3.1. Conceptualising advocacy

In 2011, Teles and Schmitt wrote that, currently, hardly any large social changes happen without some form of advocacy. Although there is a longstanding history of advocacy, going back to the anti-slavery campaigns in the 19th century, the pursuit of advocacy by NGOs has emerged on a broader scale since the 1990s, when development NGOs increasingly sought to combine advocacy with improvements to the daily lives of those living in poverty (Eade, 2002: ix; Keck and Sikkink, 1999). Currently, advocacy and its evaluation have become an individual field of theory and practice (Gardner and Brindis, 2017).

Advocacy is usually understood as supporting and actively championing a specific cause, policy goal or change. In this dissertation, I focus specifically on advocacy for development, drawing on my experience studying advocacy conducted by development organisations in the international development field. Advocacy for development[9] is defined as a 'wide range of activities conducted to influence decision makers at different levels' (Morariu and Brennan, 2009 100-108; Fagen et al., 2009). Advocacy pursues changes in social, political and organisational systems while challenging existing power structures (Gardner and Brindis, 2017; Hudson, 2001; Keck and Sikkink, 1998). In international development, change is sought towards the overall aim of pursuing a collective good that is in the public interest. NGOs combat structural causes of poverty and injustice to influence structural change (Arensman et al., 2015). The concept of advocacy for development goes beyond mere policy influence, pursuing structural changes at the international level in public and political arenas. As a body of practice, advocacy for development stands out because of its unique properties. Advocacy can be

[9] In this doctoral dissertation, I focus on advocacy for development. The case studies examine advocacy for development programmes implemented by NGOs and a transnational advocacy evaluation of these programmes. Advocacy for development is especially challenging because it includes multiple stakeholders in multiple arenas worldwide. It is thus transnational in nature, specifically pursuing goals that relate to the overarching development strategy of poverty alleviation. In this dissertation, 'advocacy' and 'advocacy for development' are used interchangeably, but always relate to advocacy for development, transnational advocacy or NGO advocacy, unless specifically stated otherwise.

conducted by a single party in a single country, but advocacy for development is often accomplished in an international arrangement, bringing together diverse stakeholders across varied contexts. This tendency, which was also observed during the MFS II ILA evaluation, brings special dynamics to advocacy for development, including specific power differentials and challenges to communication (Barrett et al., 2016; Keck and Sikkink, 1998; Wong, 2012). A broad definition therefore suits the wide range of approaches implemented under the umbrella of advocacy for development (see Figure 1).

The various strategies of advocacy for development shown in Figure 1 can be categorised as 'insider' or 'outsider' strategies (Gormley and Cymrot, 2006; Mosley, 2011). Insider strategies are about being included and creating space for inclusiveness in decision-making and policy processes. These strategies include lobbying, playing an internal advisory role in official policymaking spaces, and cooperating with targeted individuals, governments, organisations and corporations. Coalitions, social movements and networks often include multiple stakeholders working together to put an issue on the agenda, influence change or come up with solutions to existing problems. Outsider strategies, in contrast, focus on putting pressure on advocacy targets (i.e. communities, publics, policy makers, decision makers, corporations or private sector stakeholders, civil society). Pressure is applied through confrontation, such as media campaigns or naming and shaming, which is about publicly renouncing targets for their amoral conduct.

In the MFS II ILA evaluation, we saw examples of hybrid strategies that combined insider and outsider tactics. For instance, an activist media campaign to assert public pressure on decision makers turned into a multi-stakeholder strategic partnership, and an insider-focused cooperation with policy makers resulted in pressuring governments using naming and shaming and public campaigns (Arensman et al., 2015: 592). These strategies are thus not used in isolation; they may be combined to strengthen leverage, gain legitimacy, create political will to action, and influence policy changes or social transformation (Binderkrantz, 2005; Carbert, 2004; Fyall and McGuire, 2015).

Advocacy for development is complex. Figure 2 shows, in a simplified way, how advocates move across diverse contexts of organisational structures and systems (protocols, accountability relations, procedures and hierarchies) while practicing advocacy. This happens in interaction with the needs and demands of donors, targets, constituencies and partners (i.e. partnerships, social movements, networks) and in navigating these needs and demands (the *politics of results*). These donors, targets, constituents and partners also operate in their own environments with systems and structures in place. The advocate moves among these diverse arenas to get

the advocacy message across to their partners, networks, donors and targets while also conforming to the needs and demands within their organisational systems and the needs of the constituents they represent. These processes and their outcomes are not in the advocates' direct sphere of control because, although they can apply pressure and strategize for changing certain behaviour, they do not have control over the other individuals' choices. Therefore, the change processes can be said rather to be in advocates' sphere of influence. In seeking to gain influence here, advocacy for development usually stretches over large geographical areas, covering disparate languages and cultures, which makes the field even more complex.

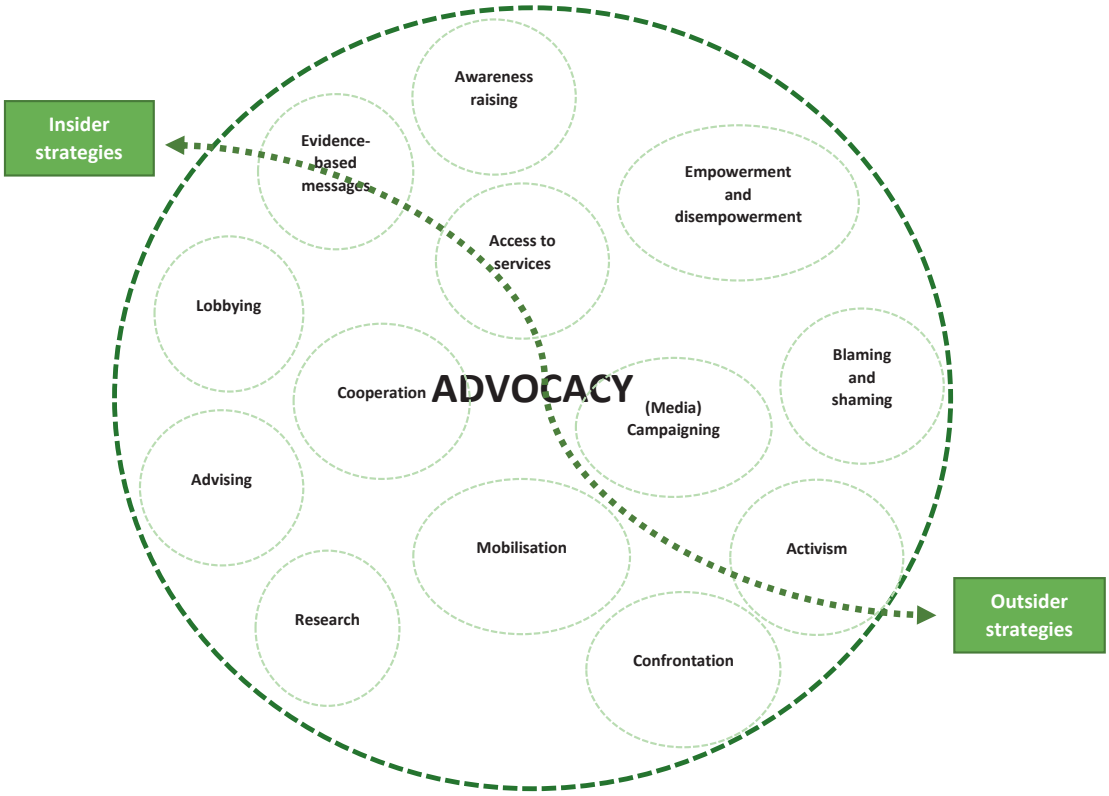

Figure 1. What is advocacy?

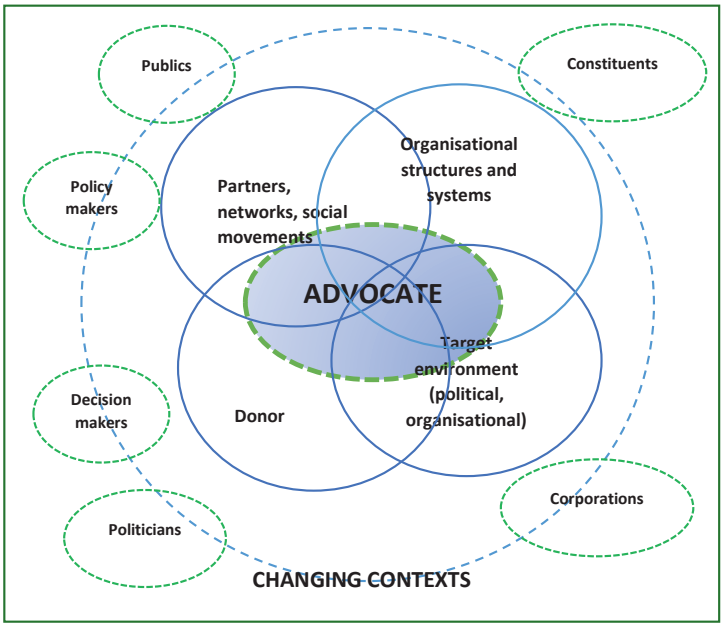

Figure 2. The advocate moves across diverse layers, levels and arenas

1.3.2. Challenges in evaluating advocacy

Why is advocacy for development difficult to evaluate? In 2005, Guthrie et al. (2005: 11) recognised that there was not yet a 'real field' or 'community of practice' on advocacy evaluation. When our team began the evaluation work in 2012, the field of advocacy evaluation was still nascent even though the contours of a field and community of practice were already visible, with leading studies by the Harvard Family Research Project (2008) and the Innovation Network (2007). A number of studies have focused on advocacy evaluation, asking 'What is it?' (Fagen et al., 2009) and 'How do we evaluate advocacy?' (Beer and Reed, 2009; Coffman, 2009; Lynn, 2014; Jones, 2011). Other studies have focused on trends, challenges and theories in advocacy evaluation (Devlin-Foltz et al., 2012; Tsui, Hearn and Young, 2014). The latest research on advocacy evaluation by Gardner and Brindis (2017) provides an overview of advocacy evaluation theory and practice, demonstrating that advocacy evaluation is currently a field of research in its own right.

A number of problematic issues with advocacy evaluation can be identified from the existing studies. Four of these problems are particularly important. First, the establishment of causal relations between the advocacy and the results is difficult. Changes are often not explicit or

visible, and most influencing happens behind closed doors or in personal interactions; the resulting changes are therefore hard to trace (Guthrie et al., 2005; Teles and Schmitt, 2011). The interim changes (Coffman, 2009) necessary for creating the conditions for change in policies or practices are especially likely to be invisible. Advocacy outcomes can include relations built, trust gained, space created for civil society voices, and the achievement of raised awareness (Arensman et al., 2015). Second, advocacy often involves many actors and factors, including multiple civil society organisations and governments working together as well as separately on the issues involved. Third, the contexts in which advocates operate are therefore usually far from stable, and the targets advocates try to influence (e.g. individuals, decision makers, policy makers, publics, societies, communities, private sector actors, other civil society stakeholders) are moving, developing, interacting and changing. Fourth, these change processes involve multiple layers, organisational structures and systems, and (power) relations at multiple levels (see Figure 2).

The main problem with establishing advocacy effectiveness is therefore that the effects of advocacy for development may be found amongst numerous other causal strands (Gardner and Brindis, 2017; Hudson, 2002; Lowery, 2013; Rogers, 2008). It can thus be extremely difficult to identify changes or outcomes. In addition, it is difficult to identify evidence of change or of contribution to the changes achieved (Chapman and Wameyo, 2001; Lowery, 2013; Rogers, 2008).

1.3.3. Conceptualising advocacy effectiveness

Clearly, the meaning of advocacy effectiveness is important for advocacy evaluation. However, advocacy effectiveness is not straightforward. Many forces are at play that make or break an advocacy effort, shaping its effectiveness. Advocacy for development happens across state lines and on multiple levels (local, national, regional, global), crossing diverse systems (organisational, institutional, governmental) and power dynamics (donor–grantee relations, target–advocate relations, emerging actors and factors, political will and space). Advocates may act on an opportunity or be successful using a certain strategy in one case but not in the next. Their efforts may be countered by powerful adversaries or emerging threats. The political space in which to act or the interest in the issue at stake can grow or shrink along with shifts in power (e.g. elections) or changing circumstances (e.g. the collapse of clothing factories in South Asia that provided momentum for clean and fair clothes campaigns in the Netherlands).

Few topics have gained as much attention as organisational effectiveness in the non-profit sector. However, this attention has not resulted in conceptual or empirical consensus as to what constitutes effectiveness or how it is best measured or evaluated. The most obvious question asked when discussing organisational effectiveness is to what extent an organisation reaches its goals (Herman and Renz, 1999). Herman and Renz (1997) have argued that much of organisational performance theory was developed from this main question, based on the assumption that organisations have goals and that these goals are measurable. Scholars of organisational performance seek to explain effectiveness by focusing on strategic management or board performance (Brown, 2005; Herman and Renz, 1997) or on organisational factors and the relation to their environment (partnerships, networks) (Herman and Renz, 1997; Provan and Lemaire, 2012). In addition, social theory scholars stress the importance of the skills of individuals to use the available resources as they see fit in their everyday practices; these skills inform the activities, decision making and strategies in organisations (Bourdieu, 1990; Emirbayer and Johnson, 2008).

Many organisation studies that focus on effectiveness have emphasised performance, without explaining how meaning is given to effectiveness or how the human element—or the interactive and strategizing nature of human actors—plays an important role. This stressing of performance (i.e. outcome-focused) is also visible in studies that discuss advocacy effectiveness (Beer and Reed, 2009; Coffman 2009; 2011; Coffman and Reed, 2009). In contrast, Teles and Schmitt (2011) have taken up the human element of the advocate and the advocate's skills in pursuing effectiveness. Although effectiveness is about the successful achievement of desired goals, as was explained above, advocacy presents challenges in this area because making change visible and relating the changes to the advocacy actions are rarely straightforward processes. Looking at the existing literature on advocacy processes, practices and evaluation, effectiveness is a main point of concern.

Table 2 summarises some of the main aspects related to advocacy effectiveness in the existing body of literature, grouped into five overarching dimensions. Advocacy effectiveness is often explained by focusing on the advocacy content and message (e.g. whether it resonates, evidence-based content), knowledge about the landscape (i.e. knowing who, what, when, how), knowledge about and use of the network (e.g. outreach, density, embedded message, balancing unity/diversity), capabilities involving an enabling external environment (relationships, seeing and being seen, dealing with demands and needs) and capabilities involving an enabling internal environment (commitment, intuitional support, capacity and skills, adaptability).

Table 2. The diverse dimensions in understanding advocacy effectiveness

Message	Landscape	Network	Capabilities involving the external environment	Capabilities involving the internal environment (enabling organisation)
Not spreading too thin, focusing on a small number of core issues (Almog-Bar and Schmid, 2013; Keck and Sikkink, 1998; Wong, 2012)	Knowing who, what, when, why and how (Coffman, 2007; James, 2011)	Network density and information (Keck and Sikkink, 1998, 1999)	Responsive to capacity needs and demands, knowing and combining diverse strategies (Almog-Bar and Schmid, 2013)	Organisational capacity, structure, strength and size (Alliance for Justice, 2015; Shawki, 2011); capacity and skills of advocates (Teles and Schmitt, 2011), including adaptability (Tsui et al, 2014; Morariu and Brennan, 2009)
Message resonance (Keck and Sikkink, 1998; Shawki, 2011)	Stakeholder analysis (knowing your players, remaining up to date); right message to the right person/institution (Sloot and Gaanderse, 2010)	Political responsibility, conceptualising power relations (Jordan and van Tuijl, 2000)	Scaling impact by expanding the scope of outreach through networks, partnerships; seeing and being seen/acknowledged at all levels	Internal work planning, focus, strategies and evaluation but remaining well prepared for the unforeseen (Guthrie, Louie and Foster, 2005)
Political salience (Wong, 2012)	Knowing the context (analysis); low profile/behind the scenes or confrontational; knowing what message is needed and how; ongoing engagement, building personal relationships with targets (Rees, 2001)	Unity/diversity balance (Saz-Caranza, 2012; Wong, 2012)	Dealing with power and with vested interests, knowing how to exert pressure (insider/outside strategies) (Gormley and Cymrot, 2006; Mosley, 2011)	Planning, monitoring and evaluation in place (Beer and Reed, 2009; Coffman, 2007); mutual learning (Leeuwis and Ringsing, 2008); open communication lines and shared responsibilities (Beer and Reed, 2009)
Well-formulated content message and content as a bridge-builder (Carpenter, 2007; Lynn, 2014)	Remaining up to date on the institutional landscape; (informal) information flows bind network members (Keck and Sikkink, 1999)	Knowledge management: knowing what to share and what not to share (Tsui, Hearn and Young, 2014)	Relationships with multiple stakeholders balanced and prioritised on the basis of values (Hudson, 2001); mobilised and responsive co-advocates/partners	Institutional support and organisational (long-term) commitment to the issue (Skene-Pratt, 2013)
Framing the message properly (positive/negative) (Corner and Hahn, 2009)		Competent ability to mobilise members (Tsui, Hearn and Young, 2014)	Accountability priorities, problem definition and defined stakeholder interest (Brown, Ebrahim and Batliwala, 2012)	Organisational centralisation of authority versus flexibility/reflectivity (Almog-Bar and Schmid, 2013; Tsui et al, 2014)

Table 2 illustrates how the existing body of literature provides these important insights into what constitutes advocacy effectiveness and what factors play a role in being effective. This

shows that there is room for complexity in understanding advocacy effectiveness, as the literature acknowledges the dynamics of relations, partnerships, and networks using knowledge, power and organisational capacity. Although this existing work does provide valuable lessons and important insight into the complex character of advocacy effectiveness, this work seems to have missed the point that effectiveness is not a self-evident qualification. Advocacy outcomes can be very different from what advocates set out to pursue in their objectives. The dynamics surrounding which effectiveness is pursued, negotiated, understood and interpreted and what this means for advocacy effectiveness and evaluation have not yet been unpacked. How advocacy effectiveness is given meaning and negotiated (what I call the *politics of results*) remains undiscussed. Questions about what constitutes the meaning of effectiveness, especially regarding the dynamic character of advocacy and the contexts in which advocates and evaluators operate are usually not included.

1.4. Research questions and objective

In my experience as both an evaluator and a researcher of advocacy for development, I found that there is a gap between how advocacy effectiveness is understood in the existing methods on advocacy evaluation and how it plays out in practice. Establishing, assessing and measuring effectiveness is a political process of identifying, measuring and presenting outcomes. In this process, there is space for manoeuvre and for interpretation when it comes to what constitutes an outcome. Existing evaluation methods are results-based, narrowly focusing on outcomes as measurable and evaluation as the objective instrument to establish these outcomes. In my role as a researcher and evaluator, I found that these methods do not provide room for understanding the political dynamics that affect the interpretation and measurement of results. The aim of this doctoral dissertation is therefore to contribute to a better understanding of the practices of advocacy and its evaluation, answering the following overarching research question:

How does the practice of advocacy for development give meaning to effectiveness, and how do the politics of results play out in the evaluation of advocacy?

This broad research question inspired a number of specific research questions:

1. To what extent are ownership and effectiveness pursued by a transnational advocacy network in tension with each other? What does this mean for advocacy effectiveness? (Chapter 2)

2. How is effectiveness negotiated, and how do interactions in advocacy evaluation give meaning to advocacy effectiveness? What implications do these processes have on evaluation? (Chapter 3)
3. What is the relation between advocacy processes and practices and the theory-driven methods used for advocacy evaluation? Does a ToC approach to advocacy evaluation do justice to advocacy processes? (Chapter 4)
4. How are advocacy outcomes identified in advocacy evaluation, and what are the challenges of and lessons learned from this process? (Chapter 5)
5. What do the findings of the above questions mean for advocacy evaluation? (Chapters 2, 3, 4 and 5)

1.5. Theoretical and analytical framework

1.5.1. Conceptualising the politics of results

Conceptualisations of effectiveness have become ensnared in an understanding of results as visible and tangible, in line with funding agreements and as defined in accountability procedures (Best, 2017; Riddell, 2014). Within this framework, the politics of results can simply be understood as the pressure put on organisations to make their often predefined and predictable results visible (Eyben et al., 2015). Politics, then, is about authoritative decision making, unequal power relations, and efforts to oppose and criticise those in public and private arenas through formal and informal activities. However, I argue that the politics of results is about more than the pressure for results. To conceptualise the *politics of results*, I highlight the processes of meaning-making that define what results, effectiveness and advocacy are about.

According to Calhoun and McGowan discussing Arendt's meaning of politics, politics arises from human plurality (Calhoun and McGowan, 1997: 1; Arendt, 1998; Schaap, 2007). This means that human beings act and start new interactions with emergent and unpredictable results (Canovan introduction to Arendt, 1998: viii–ix). Further, Arendt conceptualises politics as central to human action for questions of meaning, identity and value, giving space to the relation between thought and action (Arendt, 1998; Calhoun and McGowan, 1997; Schaap, 2007). Dahl (1964: 6) also identifies human interaction as key to politics, defining politics as 'any persistent pattern of human relationships that involves power, rule or authority.' Mouffe (2005) specifies politics as the practices, discourses and

‘rules of the game’ required for social order, and stresses that the political stems from the issues and struggles between social groups over power and resources. Mouffe (2005: 17–18) goes on to argue that, ‘while politics is not everything, everything is political’. These ideas about politics and ‘the political’ demonstrate the multifaceted nature of these concepts, which involves power distribution and values as well as decision making and negotiation. In addition, Denton and Woodward (1998: 2) emphasise that ‘communication is the vehicle for political thought, debate, and action’ and that it does not matter how politics is defined as long as it is understood that human interactions are central.

In line with these theories, I argue in this dissertation that human interactions define the meaning of effectiveness in advocacy and its evaluation (Arensman et al., 2017b; Teles and Schmitt, 2011). Building on Kerkvliet’s (2009: 232–233, 240) argument that politics shapes everyday lives, practices and behaviours, expanding the conventional forms of politics, I understand the politics of results as part of the everyday politics of practitioners and advocates in the development field. According to Kerkvliet, everyday politics involves the intrinsic human processes through which people embrace, confront, contest, comply and adjust power, authority, norms and values; these everyday interactions are often not regarded as political even though they are (Kerkvliet, 2009). The politics of results can be considered the production of meaning given to the outcomes and effectiveness of advocacy, along with the necessity of conforming to the rules of the game (i.e. the procedures demanded by the rules and regulations, accountability, and funding agreements) and organisational systems.

Key to the politics of results is the pursuit of legitimacy and credibility for the intervention or organisation vis-à-vis other stakeholders. This involves vested interests around the demands and the need to show effectiveness. This process of defining effectiveness and outcomes is profoundly political. The politics of results is the negotiated process through which outcomes and effectiveness are given meaning in the practice of advocacy for development. The interaction between needs and demands influences the underlying dynamics in the process assigning meaning to effectiveness. Because the outcomes and effectiveness of advocacy for development are difficult to operationalise, it turns out in practice that the politics of results often boils down to the question of whether the intervention or organisation succeeded in achieving legitimacy and credibility in the eyes of other stakeholders. In these processes, advocates do not only pursue results; they also interpret and negotiate the meaning of these results. The politics of results involves human interactions on the basis of which strategic considerations and decisions are made while embracing, negotiating, interpreting, adjusting and contesting procedures, norms and rules

in the quest for effectiveness. Being aware of these political dimensions makes us conscious of how they affect daily life and practice.

1.5.2. Results-based evaluation traditions in international development

The growing concerns about effectiveness put more pressure on NGOs and donors to make development results visible. This has inspired evaluations based on the idea that results can be predicted and controlled. This triggered the assumption that evaluation is or can be value-free (Marchal et al., 2012). Furthermore, it inspired the quest for general patterns of cause–effect relations (Armytage, 2011; Riddell, 2014). The increasing challenges to make effectiveness of development visible have resulted in an emphasis on performance and results-based regulatory frameworks, including evaluations (Best, 2017; Guljarani, 2011; Riddell, 2014). Central to results-based ways of working is the focus on replicable short-term outcomes. These outcomes are often required in regulatory frameworks and aligned with the intended objectives as specified in funding agreements (Guljarani, 2011; Riddell, 2014). This influences the way accountability is pursued. Accountability-driven tools and methods often work with predefined indicators to demonstrate effectiveness through measurable and tangible results (Alexander, Brudney and Yang 2010; Eyben et al., 2015), making the establishment of causal relations between activities, output and outcomes in measurable results central to the effectiveness discourse. Effectiveness then becomes instrumental (tool-driven) and political (results-driven) in nature.

The increasing demand for results-based working in evaluation developed from a rational-realist perspective, leading to an increase of theory-based and objectivity-based evaluation in the (post-) positivist tradition (Lam, 2002). Results-based evaluation is part of the (post-) positivist tradition and affects how meaning is given to results and thus to effectiveness (i.e. the *politics of results*). [10] Positivists assume an independent relationship between the evaluator and

[10] Results-based evaluation is inspired by results-based management as part of the New Public Management paradigm. The introduction of New Public Management in the 1980s brought in a results-based way of working in the development field. This came as a response to growing public pressure driven by dissatisfaction with public budgets, governments' performance and excessive public spending (see Brinkerhof, 2008). NGOs came under pressure specifically because the public trust in their effectiveness declined. To rehabilitate this trust, international development had to restructure and change, inspired by demands for more efficiency, effectiveness and accountability (Best, 2017; Brinkerhof, 2008; Sjostedt, 2017). Results-based management, with its roots in the public sector, seeks to use market strategies to improve how agencies operate to achieve better results (Hatton and Schroeder, 2007; Hulme, 2010). This includes the importance of defining expected results, using results chains, and conducting monitoring, reporting and evaluation of the achieved results (Hatton and Schroeder, 2007). At present, it is no longer possible in the public and non-profit sectors to leave effectiveness or results out of the discussion. This is integrated in organisations' overall everyday practice, as NGOs need to include performance

the evaluated, whereas post-positivists pursue objectivity but accept that contexts and mechanisms influence outcomes (Mansoor, 2003; Van der Knaap, 2004). In this tradition, realist evaluation is popular and goes beyond merely focusing on the success or failure of interventions; it acknowledges that events are rarely the result of a single causal mechanism and that outcomes are not independent of context (outcomes are context + mechanism) (Pawson and Tilly, 1997; Sayer, 2000).

The realist approach in the (post-) positivist tradition aims to systematically track the causal relations between mechanisms and outcomes. This way, realists seek both explanation for effectiveness and control in terms of predictability. This implies that a causal relation can be objectively defined and established. Porter (2015) criticizes realist evaluation by arguing that this approach to outcomes ignores agency as it also ignores the interests involved in human interactions. From a realist evaluation approach questions like: for whom are which outcomes desirable or undesirable? What are the effects of the interventions on those affected by them? What are the effects of the evaluation outcomes for those involved?; are not a point of focus. The realist approach to evaluation does not consider the role of the evaluator or the politics of the interactions between the evaluator and the evaluated in shaping the evaluation outcomes. Rather, evaluation and evidence are considered neutral and rational, failing to acknowledge their social and political construction.

The (post-)positivist traditions (i.e. realist approach, results-based) in evaluation are characterised by a search for the truth. However, not everything can be measured in the hard sense of the word, and truth is not objective per se. Increasingly, experts such as Eyben (2015, 2012, 2010), Riddell (2007, 2014) and Wallace et al. (2006) have pointed out that making effectiveness measurable carries a risk because it leads to a narrow focus on those results that are visible. Riddell (2007: 12) has argued that this narrow focus prioritises short-term, tangible results, while paying less attention to addressing significant, underlying, long-term development problems. Eyben (2012) added a call for recognising that not everything in international development is measurable in terms of visible and rigorous numbers: '[A] dialogue about the change desired and a theory of how to achieve it can be useful provided we recognize that any explanation is partial, contingent on context and needing to be regularly checked against reality, as experienced from diverse perspectives'[11] (Eyben, 2012).

information in budgets, quarterly and annual reports, evaluations and management plans (Eyben, 2015; Mayne, 2007; Wallace, 2006).

[11] Eyben, 2012. Online discussion. 11 March 2012. 'ToC: breaking out of the results agenda', see: http://bigpushforward.net/archives/1419

Complexities need to be understood and recognised (Ramalingam et al., 2008). Eyben (2012) highlighted that, no matter how we aim to plan, manage and control the world around us, we always need to understand what we do, how we do it, what it leads to and what it means in relation to the contexts, the diverse realities we live in and the experiences we have, including the diverse interpretations that are at play.

Constructionists therefore emphasise that the (post-)positivist perspective is too simplistic and does not do justice to the complex and dynamic nature of reality. They argue that there can be no generalisable or universal claim of truth or reality (Van der Knaap, 2004). Rather, Eyben (2015) and Wallace (2006) stress that regulatory frameworks are defined by those in power. Those in power demand results from those who must demonstrate these results for organisational and professional survival. This implies a power imbalance. Approaching evaluation from a perspective of objectivity assumes evaluation roles, concepts, purposes, methods and outcomes are non-negotiable. Hence, this does not allow for challenges to the existing power relations, as agency and interests involved in evaluation are not taken into account. A (post-)positivist approach to evaluation falsely implies an ease of objectively evaluating and delivering outcomes, although, in reality, these things are much more political in character. Taking such an approach to evaluation risks overlooking how effectiveness is given meaning as part of the *politics of results*, which includes underlying dynamics such as pursuing, negotiating and interpreting results. Evaluation is political in terms of decisions about which outcomes are included, which outcomes define effectiveness, and who decides what is measured, how and why.

1.5.3. Meaning-making in advocacy evaluation and the politics of results

Inspired by the criticism of results-based evaluation, the challenges faced in evaluating advocacy, the underlying politics involved and the need to surface how advocacy effectiveness is given meaning in evaluation, I take a social constructivist approach, using practice-based theory. An actor-oriented perspective on effectiveness (Giddens, 1984; Long, 1989, 2003) considers meaning to be made through the process of interactions. Individuals are considered actors with *agency* who make their own decisions based on beliefs and values. To make this more explicit, I follow the tradition of practice-based theory, in which individuals are considered strategic actors and strategy is considered something an actor does, decides and knows in interaction with other actors and contexts (Jarzabkowski, 2005).

Evaluating advocacy effectiveness challenges how effectiveness is given meaning. Results are central to the evaluation of effectiveness, but the realist and (post-)positivist traditions in evaluation tend to simplify complex dynamics into predictable and controllable portions of information that assume complex reality can be captured in an objective, rigorous and unbiased manner (Guljarani, 2011: 207). Lo and Mueller (2010) call this 'physics envy', which describes the quantitative aspirations of social sciences building on the belief that it is possible to build predictive models (Lo and Mueller, 2010; Wilson and Thomas, 2011). Adding to this is Hulme's (2010) perspective that human development—as development is intrinsically a human-driven process—and results-based management are two 'unlikely bedfellows'. Advocacy, as demonstrated above, is a human-driven process, and its results are often uncontrollable, ambiguous or not self-evident. This causes tensions between the way effectiveness is evaluated and the difficulty of measuring effectiveness in advocacy.

Existing evaluation methods do not give room to meaning-making, as they are embedded in the realist and (post-)positivist traditions that falsely imply objectivity in evaluating and delivering outcomes. Guba and Lincoln (1989: 7) have previously argued that evaluation is not necessarily objective: 'To approach evaluation scientifically is to miss completely its fundamental social, political and value-oriented character.' Lowery (2013) has emphasised the political nature of advocacy evaluation. The role of an evaluator is therefore more than a gatherer and assessor of data; an evaluator is also a negotiator and facilitator, managing and navigating learning, expectations and interests (Markievicz, 2005; McDonald, 2008; Sharkey and Sharples, 2008; Taylor and Balloch, 2005). Teles and Schmitt (2011:43) underscored this point, arguing that 'it is far better to acknowledge that tacit knowledge and situational judgment are what really underlie good advocacy evaluation, and to find evaluators who can exercise that judgment well. It's the evaluator, rather than the formal qualities of the evaluation, that matters.' These scholars stressed that advocacy evaluators need to deal with sources of information that are often biased. Evaluation—and advocacy evaluation in particular because of its complex and dynamic character—should thus take into account the politics that are involved in meaning-making around effectiveness.

Dedicating attention to meaning-making around advocacy effectiveness is necessary to surface the dynamics of effectiveness and is all about the politics of results. Advocates operate in complex and often highly political environments where situations can change quickly, and many actors and factors play a role in shaping these contexts. In such insecure or unplannable situations, advocacy effectiveness is highly political and dependent on the nature and quality of advocates' practical judgements and *fingerspitzengefühl*. The question of when advocacy is

effective is not easily answered, because the meaning of advocacy effectiveness is embedded in political dynamics such as negotiations about what constitutes effectiveness and how outcomes are reported or identified.

Evaluation is political (i.e. who decides what effectiveness is, how and why?), and advocacy in particular presents challenges to the current traditions of evaluation (realist and positivist/post-positivist). In this doctoral dissertation, I will demonstrate that evaluation happens through interactions, including confrontations and discussions about how to position particular outcomes vis-à-vis others. This involves politics in tension concerning the need to show results (accountability), the need to achieve results (credibility and legitimacy), and everyday practices in which results are not necessarily tangible, visible or even measurable. The *politics of results* is the production of meaning given to results in the quest for pursuing, negotiating, interpreting and achieving effectiveness.

From a social constructivist and actor-oriented perspective, it is possible to do more justice to advocacy dynamics and the politics involved, because this provides space to understand how meaning is made and given to effectiveness. 'Zooming in' on interactions between advocates and evaluators, it becomes clear that these practitioners act strategically and with agency to make practical judgements as they manoeuvre, negotiate, act, react and adapt to (changing) circumstances. Studying these processes and practices from a social constructivist perspective gives an understanding of advocacy as inherently human-driven, putting interactions at the centre of meaning-making. Rather than asking what was achieved (results focus), the social constructivist and actor-oriented perspective provides the analytical lens and practical approach to ask the questions that open up the so-called 'black box' of strategic practices to really understand the meaning of what was undertaken, why, how, by whom, with whom, to what end, and with what strategic decisions and considerations.

1.6. Methodology

Being a researcher and evaluator in one of the largest evaluations of transnational advocacy for development worldwide[12] gave me a unique opportunity to study and analyse advocacy and its evaluation. The complexity of advocacy challenges evaluators to comprehend its multilevel, multi-sited, multifaceted and dynamic nature, and puts conventional evaluation methods to the test. During this evaluation, we (the evaluators) were faced with complications of political

[12] Dutch Co-financing System second term, or, in Dutch, *Medefinancieringsstelsel tweede termijn* (2010–2015) Evaluation of International Lobbying and Advocacy.

dynamics in establishing effectiveness and identifying outcomes. The intention of my PhD was therefore to contribute to a better understanding of how effectiveness is given meaning in advocacy for development practices and how the politics of results play out in the evaluation of these efforts.

1.6.1. An iterative approach

The evaluation's aim to focus on results, learning and analysis provided the space to learn about advocacy and to develop advocacy evaluation methodology. The evaluation sought to merge two positions in evaluations that are sometimes positioned as irreconcilable: the linear evaluation tradition, which compares outcomes at a certain moment in time against a set of objectives, and the evaluation perspective that acknowledges complexity and recognises the adaptive capacities of programmes. By reconstructing programmes' ToCs, we sought to establish a baseline of objectives, strategies and assumptions. Taking a ToC approach was also intended to identify changes and to acknowledge the complex everyday realities in advocacy programmes. Using multiple types of data sources (document study, interviews, observation, discussions and workshops), the evaluation aimed to account for the interaction between the internal and external processes that shape eventual outcomes. This provided space for understanding the agency of actors and the adaptive capacity of their institutions.

The evaluation was set up to be a collective effort, with a team of diverse evaluators working in cooperation with the evaluated programme staff. In the first month, we organised a workshop with staff members of the evaluated programmes to establish common ground on what to evaluate and how to evaluate their programmes. This was based on the assumption that we needed to pick their brains to understand what we were evaluating and what challenges we were going to encounter, and to find common ground on the solutions.

For the evaluation, we used an iterative approach. We designed the methodology progressively as we learned about advocacy during the process from the evaluated programme staff and while gathering data. Collectively with the evaluated programme staff, we discussed and negotiated outcome descriptions and outcome claims, which were substantiated using data external to the programmes. Our analysis and interpretations of the findings were part of the meaning-making regarding the data gathered, discussions within the evaluation team and discussions with the evaluated programme staff. This was not a straightforward process. As evaluators, we went back and forth discussing important evaluation challenges and methodology development within the evaluation team. In addition, we continually discussed

the evaluation with staff members of the evaluated programmes. This included negotiations on the terms of reference of the evaluation, conversations about the evaluation questions and their interpretations, and discussions about the methodological choices for the evaluation. Discussing these issues in interactions with those who were being evaluated was also meant to contribute to learning on both sides: learning for us as evaluators and learning together with those being evaluated.

1.6.2. Methodological design and data gathering

This dissertation presents the analysis of a number of challenges that were encountered during the evaluation. These challenges were researched and analysed in depth based on my experiences as a researcher and evaluator. The dissertation is thus not an account of the entire evaluation; rather, it focuses on a number of specific problematic issues that arose during the evaluation. These issues were the starting point for in depth-research and further analysis, and the different chapters present the results of this exploration.

The study partly had the character of a multi-sited ethnography, which refers to an ethnographic study following actors, programmes or processes through time and space (see Marcus, 1995). Marcus has highlighted that 'multi-sited research is designed around chains, paths, threads, conjunctions, or juxtapositions of locations in which the ethnographer establishes some form of literal, physical presence, with an explicit, posited logic of association or connection among sites that in fact defines the argument of the ethnography' (Marcus, 1995: 105). Multi-sited research therefore suits the study of advocacy for development evaluation. This type of advocacy is usually transnational. Multi-sited ethnography is meant to make sense of the rapidly globalising world, where actors operate beyond borders and beyond their own social, cultural or political spaces. This approach also suits the analysis of advocacy for development in that it involves multiple organisational levels and structures of accountability between the stakeholders involved in an evaluation. For this research, I followed three advocacy programmes across time (2012–2015), geographical space (the Netherlands, the United States [New York], Ghana, Ethiopia, the Philippines, Belgium [Brussels], Kenya, Mozambique, Uganda, Burundi and Colombia) and organisational space (networks, partnerships, and internal organisational levels and systems). In this way, I was able to capture four levels of multi-sitedness:

1. The first site is the case study level: This allows for broader as well as deeper insights into multilevel advocacy for development through surfacing an overall understanding of the processes involved.
2. The second site relates to the transnational character of the cases: Some operate through a global secretariat in the Netherlands or a regional office in Ethiopia, but all of the explored cases are transnational and work in a networked manner. Following the processes thus meant following the advocacy across borders and through key entry-points, seeking the links across advocacy at the global, regional, national and local levels. I followed the processes through the (participatory) observation of internal and external meetings, as well as public events and advocacy actions or events. These sites were chosen independently (but always in cooperation with the organisations involved) as characteristic of and key to the processes I sought to understand.
3. The third site involves the multilevel character of the advocacy processes I studied: I sought to understand and unravel advocacy work with multiple stakeholders in various arenas, considering the meaning of the networked advocacy, the aspect of representation (for whom? how is the agenda set?) and the relation to effectiveness.
4. The fourth site of this multi-sited ethnographic study is the reflexive study of the evaluation process through time (2012–2015) and across space, in local, national, regional and global contexts.

The everyday interactions, practices and politics involved in the evaluation served as the point of entry, providing ethnographic outcomes that can be understood as co-created and negotiated between the evaluators and the evaluated. Hilhorst (2003: 83) has argued that the analysis of interfaces (i.e. interactions where actors wield and yield power) can provide insight into how goals, perceptions, interests and relationships are shaped and reshaped as a result of interactions. Studying these interfaces means examining meaning-making, where actors have agency over their decisions and are central to how meaning is given to processes, procedures and concepts. Meaning is made through communication, co-creation, negotiation and interpretation (Hilhorst, 2003; Kerkvliet, 2009; Yanow, 1993) occurring at these interfaces of interactions between evaluators and those being evaluated. This approach situates this research in the constructivist epistemological tradition.

I used different types of qualitative data collection methods: (participant) observation, semi-structured face-to-face interviews, formal and informal meetings and discussions, and the collection of primary and secondary documents (internal organisational documents, notes,

briefs, media outlets, websites, reports, etc.). The initial interviewees and contact persons were from the advocacy programmes under evaluation, and snowball sampling was then used to solicit new contacts through these existing contacts. These new contacts were people who were directly or indirectly involved in the advocacy programmes or were otherwise relevant for the advocacy, such as advocacy targets, who were knowledgeable about the thematic areas or specific arenas.

Doing multi-sited ethnographic research enabled me to observe advocates while they were doing advocacy. Participant observation ranged from passive observation of diverse advocacy processes and interactions to active participant observation as an evaluator in the evaluation. Observations were conducted in internal team meetings of the advocacy programmes and of the evaluation team, lobby meetings, conferences and (public) events, formal and informal discussions, partner meetings and strategy meetings. As I was both a researcher and an evaluator, the interactions I observed and recorded also required introspection from me from the standpoints of both of these roles. I also conducted interviews.[13] Some interviews were recorded in notes, and others were audio recorded. It was not possible to audio record all of the interview sessions because of sensitive issues being discussed or because the location of the interview was not amenable to recording. These interviews ranged from 30 minutes to three hours. I interviewed advocates in the diverse spaces where they operated, as well as other stakeholders relevant to the advocacy processes, observing how the advocates acted and interacted when changes occurred and how they interacted with other stakeholders. This meant that it was possible to observe, study and interview a diverse set of advocates working in the evaluated programmes in the varied settings in which they worked. This gave me the opportunity to see advocates in action during meetings, while having drinks or in hallways between meetings, thus gaining a sense of the specifics and the meaning behind what they were doing and how they acted strategically.

The primary and secondary documents provided important insight into advocacy processes and practices because these documents reflected discussions and decision making on strategic directions, relations with partners or targets, and influence wielded and translated into documentation such as email conversations or tracked changes within documents. The secondary data sources included media outlets, reports, policy briefs, workshop documentation and event outreach. These data sources provided contextualisation, but also insight into the uptake and outreach of the advocacy.

[13] See Annex 2 in this thesis for an overview of the interviews and meetings observed.

1.6.3. Wearing different hats

The research on the MFS II ILA evaluation was part of my PhD study. I was the sole full-time evaluator and researcher included in the process, researching advocacy and advocacy evaluation and also ethnographically studying the evaluation process as a whole. This meant double positionality—being an evaluator and being a researcher, following, observing and studying the evaluation as a multi-sited ethnographic study. As the only researcher on the evaluation team who was employed full-time, I had the opportunity to be involved in the evaluation at multiple levels: as an evaluator, part of the management team, a thematic cluster coordinator, a researcher studying advocacy and advocacy evaluation, and a researcher and observer of the evaluation processes. This dissertation is therefore an in-depth analysis of and a reflection on the evaluation process as a whole, zooming in on key challenges and lessons learned with regard to evaluating complex and multi-layered processes such as advocacy.

Part of this dual position is the focus of my research on meaning-making in evaluation. As a researcher and analyst, besides being an evaluator, I had the access necessary to observe and participate in the evaluation as a process of interactions, negotiations, and decision making. This provided a unique opportunity to analyse meaning-making as part of an evaluation process. The evaluation was a human endeavour, and, as such, human interactions and actor agency were central to shaping the discussions, negotiations and outcomes. In my dual position as evaluator and researcher, I am aware that my perspectives and findings were shaped by this setting and that my role as an evaluator influenced the processes. This dissertation therefore reflects on the lessons learned from the perspective of a researcher-analyst studying an evaluation in which she also participated as an evaluator.

My PhD project was woven into the broader advocacy evaluation, with consent from the donors and stakeholders involved (SGE, the evaluated organisations, the evaluation team, the NL MFA and NWO). This consent enabled me to conduct an ethnographic study of advocacy practice and its evaluation. Moreover, my positionality in the evaluation gave me access to privileged information on advocacy processes. In interactions and interviews, I always stressed my dual role as both an evaluator and a PhD researcher. However, ethnography is an evolving process that relies on observations, making it impossible to inform participants beforehand about the direction of my analysis. It was therefore not possible to obtain written consent in all of the interactions, but, again, it was always explained and stressed that I was operating in a dual role. Regarding publications from this research, it was agreed with the stakeholders

involved that these would be anonymised unless explicit consent was given by the organisations not to anonymise. In the chapters in this doctoral dissertation, I anonymised the individuals involved and the interviewees, as well as the quotes. I did not anonymise the organizational identity of the case studies in Chapter 2 and 3, as I received consent from the organisations involved. In addition, I provided these organisations with a chance to respond and react to the articles before publication.

1.7. Outline of this dissertation

Building on the recognised challenges around advocacy evaluation and using the lens of the *politics of results*, this dissertation addresses the diverse dimensions that make evaluating advocacy effectiveness problematic. This work is based on an analysis of a unique dataset from a multi-sited ethnographic study of a transnational advocacy evaluation (2012–2015). The four empirical chapters represent elements of the effectiveness discussions, illustrating the diverse political elements that shape how advocacy effectiveness is perceived, pursued, assessed and negotiated. The first part of the dissertation demonstrates the complex dynamics of advocacy processes, unpacking the politics that play a role and the negotiations and navigations of advocates and evaluators. The second part of the dissertation discusses in more depth the issues around advocacy evaluation, the limitations to existing methods and the problematic identification of advocacy effectiveness. Following this, the final chapter brings together an analysis of the empirical findings and theoretical discussions and concludes by looking forward to what these findings mean for future practices in advocacy evaluation.

Chapter 2 explores how ownership and effectiveness are simultaneously pursued by a transnational advocacy network, examining the tension this brings forward. This chapter is about what this tension means for how advocacy effectiveness is shaped through interactions between these two main objectives pursued by the examined network. More specifically, the chapter reports the results of an in-depth study of a transnational advocacy network that pursues effectiveness while simultaneously pursuing shared ownership within the network and within its advocacy messages and identities. In pursuing effectiveness in international development, shared ownership is assumed to be a condition for effectiveness. Academic studies question this relation, claiming shared ownership can instead lead to ineffectiveness. This chapter analyses the interplay between ownership and effectiveness on three dimensions: collective identity, the accountability process and the advocacy message. The question is not about being more or less

effective, but about the processes that shape the meaning of effectiveness in particular institutional constellations.

Chapter 3 deals with the dilemma of negotiating effectiveness in advocacy evaluation. This chapter illustrates that the establishment of advocacy effectiveness in evaluation is negotiated, dependent on interactions among the stakeholders involved. Interaction-based evaluation is therefore unavoidable. This chapter challenges the objective and rational idea of evaluation by arguing that advocacy evaluation is inherently socially and politically constructed through interactions. Understanding evaluation as a political process means that there is room for discretionary judgement and space for negotiations when it comes to deciding on methods, conceptual clarity and results. These negotiations define what effectiveness is and how it is interpreted, measured and presented. The negotiations thus contribute to the quality and accuracy of evaluating advocacy effectiveness. Focusing on interaction provides room to explore and address the multiplicities of meaning around identifying, measuring and presenting outcomes.

Chapter 4 addresses the tension between theory-driven evaluation and the practice-based needs encountered when evaluating advocacy. Assessing effectiveness is mostly done through theory-based approaches, leaving practice-based evaluation underrepresented. Currently, the ToC approach is widely used for the planning and evaluation of development programmes. This approach is considered especially suitable for complex interventions such as advocacy. This chapter discusses and questions this assumption, arguing that the ToC approach focuses on cause–effect logic and intended outcomes, failing to do justice to the recursive nature of complex interventions such as advocacy. The chapter illustrates how advocacy evolves through recursive interactions, with outcomes that are emergent rather than predictable. The chapter proposes foregrounding ‘practices of change’ by emphasising (human) interactions, using the analytical lenses of strategies-as-practice and recursiveness. This gives room to emergent outcomes and implies a different use of the ToC, making a clear distinction between the theoretical reality models and the real world of practices.

In Chapter 5, the understanding of advocacy outcomes as self-evident is questioned, following from the recognition that advocacy can be recursive and its outcomes emergent. Current advocacy evaluation methods claim to do justice to the complex and dynamic nature of advocacy (Coffman, 2011; Devlin-Foltz et al., 2012; Keystone, 2008; Mayne, 2009; Wilson-

Grau and Britt, 2012). However, in reality, these methods do not consider the identification of outcomes to be problematic in its own right. Outcomes are considered to follow self-evidently from a programme's actions. In analysing the evaluation of outcomes in a transnational advocacy evaluation, I found that advocacy outcomes are not self-evident. Rather, evaluating advocacy outcomes provide room for interpretation, the outcomes are constructed in multiple realities, and this creates space for politics to shape the identification of outcomes.

In the concluding chapter (Chapter 6), I reflect on the implications of this dissertation for research and practice and suggest directions for future research. This chapter briefly summarises the findings from the four chapters discussing the main elements of how advocacy effectiveness is given meaning in evaluation and the role played by the politics of results through negotiations, interactions, tensions between theory and practice, and politics. The chapter centres on the themes of advocacy effectiveness, meaning-making and the politics of results, which guide the overall dissertation.

2. Does local ownership bring about effectiveness? The case of a transnational advocacy network[14]

In international development, shared ownership is assumed to be a condition for effectiveness. Academic studies question this relation, claiming shared ownership can instead lead to ineffectiveness. This study analysed the interplay between ownership and effectiveness in a transnational advocacy network for conflict prevention observed 2012–2015. Building on recent discussions about balancing unity and diversity in networks, this chapter unpacks the ownership/effectiveness relationship into three dimensions: collective identity, accountability processes and a shared advocacy message. We find that the question is not about more or less effectiveness, but about the processes shaping the meaning of effectiveness in particular institutional constellations.

Keywords: ownership; transnational advocacy networks; advocacy; networks; effectiveness; development; politics of effectiveness

[14] This chapter was accepted as an article with minor revisions and published by Third World Quarterly as: Arensman, B., van Wessel, M., and Hilhorst, D. (2017). Does local ownership bring about effectiveness? The case of a transnational advocacy network. *Third World Quarterly*, *38*(6), 1310-1326.

2.1. Introduction

For three years, we followed the Global Partnership for the Prevention of Armed Conflict (GPPAC), a transnational advocacy network (TAN) active in conflict prevention and peacebuilding. GPPAC values shared ownership and seeks to make it integral to their identity and operations. GPPAC also sees shared ownership as key for achieving effectiveness. This stance creates challenges and tensions for the network as a whole, as well as for its constituent parts, raising questions such as 'Whose effectiveness?', 'What effectiveness?' and 'How does the effectiveness of the network as a whole relate to its effectiveness for individual members?' Addressing these questions contributes to our understanding of how the ambitions of ownership and effectiveness relate to each other in practice, and with what consequences. This is especially relevant considering that, in international development cooperation policy and practice, ownership is assumed critical for effectiveness (OECD, 2008), while research calls this into question (Richmond, 2012; Reich, 2006).

The increasing emphasis on effectiveness in the international development world calls for controlled and managed processes, while also demanding local ownership based on the assumption that ownership results in sustainable development and supports effectiveness (OECD, 2005, 2008). The principle of local ownership carries broad consensus, valuing home-grown solutions to conflict and locally driven partnerships. In some cases, donors explicitly require local ownership. However, the notion that shared ownership is a policy condition for development effectiveness has been challenged by past research, often because the operationalisation of ownership is ambiguous and more donor-centred than people-centred (Richmond, 2012, 354-56; Reich, 2006, 6-8). Ownership and effectiveness are therefore considered primarily political concepts, but they are widely used in development agencies.

Based on a case study of a TAN, the present work examined how organisations engage with the potential synergies and tensions involved in the simultaneous pursuit of ownership and effectiveness. Networks are considered increasingly important for sustainable development (UN, 2015: 28), and a 'hallmark of the development industry' (Henry et al, 2004: 839), presenting a potentially counter-hegemonic force questioning the structures and power dynamics in the development field. However, networks are also influenced by political processes (Katz and Anheier, 2005: 247; Hudson, 2001: 335), including political (donor-instigated) pressures for effectiveness. Especially relevant for this chapter is the argument that shared ownership can be in tension with effectiveness in advocacy networks. Recent

publications argued, for example, that when global advocacy is too accommodative of local diversity it risks to become ineffective (Wong, 2012; Saz-Carranza, 2012).

TANs are 'networks of activists, distinguishable largely by the centrality of principled ideas or values in motivating their formation' (Keck and Sikkink, 1998: 1). These networks face the fundamental tension of the network being a collective made up of diverse organisations. Recent work has articulated this in two ways. Wong analysed the 'transnational dilemma' between the centralisation/decentralisation of the network structure and its agenda-setting power (Wong, 2012: 56-59), and Saz-Carranza described the 'unity–diversity paradox' in network management (Saz-Carranza, 2012: 64-82). These studies focused on the need to find a balance between the network as a collective and the understandings, perspectives and interests of individual organisational members.

This chapter illustrates how, in relation to these tensions in networks, the pursuit of ownership shapes the drive for effectiveness. Our aim was to provide a deeper understanding of the layered, socially constructed reality of this relationship as it is expressed in organisational discourse and practice. Ownership in development is a particularly slippery concept, which lends itself better to policy rhetoric than empirical study. In this chapter, we break down the idea of ownership into the dimensions of collective identity, accountability processes and shared advocacy messages.

We begin by discussing the key debates that informed this study, presenting previous work on effectiveness, ownership and the tensions between them in networks. Next, we describe the methods and introduce the studied network. Then, the findings section demonstrates how processes of collective identity, accountability and advocacy influence the pursuit of effectiveness. The conclusion draws out broader implications for theory, practice and future research.

2.2. Effectiveness and ownership in development

Effectiveness is usually defined in instrumental and rational terms about achieving pre-defined objectives.[15] Subsequently, 'value for money' or 'results-based financing' approaches are implemented by various countries and multilateral donors.[16] Political and public criticism

[15] See the OECD DAC criteria and guidelines on evaluation (OECD DAC, *Criteria for evaluating development assistance*, 1991; 2005).
[16] See for example: ICAI, *ICAI's approach to effectiveness*; DFID, *DFID's approach to value for money*; Yogita et al., *Output-based aid*; World Bank, *Results-based financing*.

of development effectiveness has increased, and budgets and support for development are declining. In this context, increasingly, effectiveness criteria are closely linked to funding streams and are often reduced to measurable short-term results (Riddell, 2014).

At the same time, international development policy includes ownership as a condition for achieving effectiveness. In the mid-1980s, ownership was rooted in the idea of 'people-centred' development. In the 1990s, ownership became increasingly understood as 'state-centred', based on the assumption that sustainable development only works when it is country-owned (Eyben and Savage, 2013; Richmond, 2012; Korten, 1987). However, in a simultaneous and contradictory development, the demand for development effectiveness results increased the focus on technical and operational control of development processes (Eyben, 2010). To navigate these contradictory demands, inclusive or democractic ownership became framed as a form of participatory consensus-building amongst the 'recipient country's relevant actors from state and society about the content and implementation of development-enhancing' (Faust, 2010: 516). This process illustrates the continued challenges between ownership and effectiveness that remained unresolved in practice (Booth, 2012; Faust, 2010).

These challenges emerge in operationalization and implementation of ownership as part of the effectiveness agenda (Donais, 2009; Reich, 2006). Fowler stressed that, in the past, a lack of local ownership has led to aid ineffectiveness, resulting in the logical conclusion that development aid needs to be locally owned (Fowler, 2000). However, this argument has received criticism. Reich has argued that local ownership is a vision rather than a practical objective, lacking in observability and contradicted by the donor-driven agenda (Reich, 2006). Similarly, Richmond has asserted that state-centric ownership does not promote inclusive partnerships. This contradicts the idea that ownership facilitates a sense of agency and flexibility among the represented groups and individuals. The concept of ownership rather suggests being a political means for legitimation and donor requirements (Richmond, 2012), pointing towards the socially and politically constructed nature of ownership and raising questions regarding how it actually relates to effectiveness (see also: Marston, 2000).

Although ownership and effectiveness are paired in development policy, there is a contrast in approaches to the two concepts. The language on ownership acknowledges that different actors may have different agendas and priorities, whereas the language on effectiveness treats 'results' as unambiguous and rational. In reality, questions of 'Whose ownership?' should be paired with questions of 'Whose effectiveness?' (Chambers, 1995: 14). Policy documents are not explicit about the assumed relation between ownership and effectiveness; ownership and effectiveness are related, but it is not clear how they should operate together in practice (Donais,

2009; Reich, 2006). As a result, explications of just how ownership and effectiveness intertwine cannot be found in policies but must be read in the practice of development. The nature of the relationship between ownership and effectiveness becomes visible in the arrangements unfolding around resource flows, commitments, demands and conditions, and accountability among donors, recipients and constituents (Eyben et al, 2015; Riddell, 2014). Hence, questions such as 'Effectiveness for whom?', and 'When is a programme or a network effective?' (see also: Provan and Kenis, 2008; Chambers, 1995) must be answered by studying the social negotiations and emergent properties of development practice.

2.3. Effectiveness and ownership in advocacy networks

Initial research on TANs has reported tensions between effectiveness and ownership (Shawki, 2011; Hoskyns, 1999). TANs are inter-organisational, goal-directed networks consisting of diverse member organisations that pursue changes in policies, practices and behaviour (Saz-Carranza, 2012). These networks must navigate their members' diversity in terms of geographic location, cultural and social differences, capacity and resources, while also pursuing a collective purpose, identity and strategy. According to Hoskyns, the danger for networks is trading fluidity and diversity for structure and coherence, which results in prioritising some interests over others (Hoskyns, 1999: 80). Hence, tensions arise between the network as a collective and the network as constituted by individual organisations, each with their own practices, identities and goals.

Wong has characterised this tension as a 'transnational dilemma' concerning networks' structures, governance and effectiveness. In Wong's interpretation, effectiveness is the political salience organisations have for 'at least part of their agendas' (Wong, 2012: 54). Her analysis explains political salience using the organisational structure and the roles and practices of advocates, which reflect the distribution and use of power within an organisation. Comparing different networks' handling of this transnational dilemma and its consequences for effectiveness, she found that effectiveness depends on the distribution of power within the network and its use of this power externally. These aspects of power influence how agendas are proposed, set and implemented. Determining these things requires leadership, a sense of collective and the power to make decisions, all of which provide relevance and legitimacy to the network.

Saz-Carranza approached the tensions in networks by identifying a unity–diversity paradox in network management. For Saz-Carranza, effectiveness is the ability to carry out collective

action, as well as the achievement of the organisation's mission (Saz-Carranza, 2012: 57). Crucial here is the network's need to value diversity as its 'power base', which provides it with credibility and legitimacy, while the network develops the power to act collectively and translate this diversity into access and action. Hence, the unity–diversity paradox implies that, although network diversity and unity are both necessary to achieve effectiveness, they may also undermine one another (Saz-Carranza and Ospina, 2011: 30).

The two studies discussed above both argue that effectiveness is related to the distribution and use of power within networks. They maintain that how networks deal with diversity and the collective influences of organisational structure, agenda setting and power relations has a clear bearing on effectiveness. This work has shed light on the inner dynamics of networks that shape effectiveness, as well as its meaning. However, Provan et al. have stressed that the meaning of effectiveness also depends on external factors such as the sectors in which a network operates, including the environment's political dimensions (Provan et al, 2007). Therefore, for networks operating in international development, the political dimension of effectiveness in results needs to be considered.

Building on this previous work, the present study considered the meaning and practice of effectiveness in transnational advocacy to result from interactions between factors associated with the context where the network operates and internal power dynamics. It is through these interactions that the understanding of ownership and effectiveness evolves and is challenged. Thus, we examined tensions between ownership and effectiveness in the interface between network dynamics and the surrounding complex environment. As both ownership and effectiveness emerge from practice, we sought to understand the relationship between them through a case study focusing on three key variables of transnational advocacy: identity, accountability and advocacy substance.

Collective identity is well studied in relation to social movements. According to Gamson, 'any movement that seeks to sustain commitment over a period of time must make the construction of collective identity one of its most central tasks' (Gamson, 1997: 27). Collective identity is essential for network effectiveness because it is considered important for mobilisation, participation and sustaining commitment (Hunt and Benford, 2004; Jasper and Polletta, 2001). As transnational advocacy is also done in networks, we maintain that collective identity is equally important for TANs. Collective identity concerns shared ideas about belonging to the network and what the network stands for, as well as its image and representation.

Accountability processes are crucial for questions of ownership. Accountability can be understood as a process of being held responsible and taking responsibility for actions and performance at multiple levels (Ebrahim, 2003a + b). Networks are accountable to multiple actors, including members, donors, constituents, partners and the communities they claim to represent (Ebrahim, 2003a). In the case of shared ownership, accountability is thus mutual and multi-layered. The idea of ownership means, in the first place, that agencies are accountable to those for whom development is meant (Richmond, 2012). Because ownership is a key objective in the donor community, organisations also need to account for their ownership practices to donors and other stakeholders.

The advocacy message conveys the strategy for ensuring effectiveness. This concerns who and what the TAN seeks to influence and how this will be achieved. According to Jordan and Van Tuijl, NGO advocacy is 'an act of organizing the strategic use of information to democratize unequal power relations' (Jordan and Van Tuijl, 2000: 2052). This means that the message is more than a tool to influence policy; it is also a way to renegotiate which voices are heard, how diversity in expertise is valued and how ownership over development is established.

2.4. Methods

Data were collected by following GPPAC from 2012 to 2015, focusing specifically on the advocacy component of their peacebuilding work. GPPAC was subject to an evaluation of international advocacy programmes financed by the Netherlands Ministry of Foreign Affairs, in which the researchers were involved.[17] The evaluation concerned eight advocacy programmes. We selected GPPAC as the case for this study because of its explicit co-prioritising of effectiveness and ownership. The researchers were never insiders in the network, but interacted in dialogue with network staff and members. We observed and reflected on the processes as outsiders, with the primary goal of evaluating while also seeking a deeper understanding of networked advocacy processes.

The data were gathered through multi-sited ethnography, an interdisciplinary method developed to study the interconnections in a globalising world order while examining 'dichotomies such as the "local" and the "global", the "lifeworld" and the "system"' (see: Marcus, 1995: 95). Following the network's advocacy programme, we examined an array of

[17] See for more information on the MFS II ILA Evaluation the extensive summary by Barrett et al., *Advocacy for development* (2016); or the extensive evaluation report by Arensman et al., *MFS II Joint Evaluation (*2015).

transnational advocacy. We focused on the interplay between intra-network processes and external interactions.

We gathered data through participatory observation, document analysis and interviews with stakeholders inside and outside the network. The data collection process lasted three years and was done in a variety of spaces, including extensive interactions with actors in the network's Netherlands headquarters and during brief periods of fieldwork in Ghana, the Philippines and the United States (New York). We conducted 95 semi-structured interviews and formal and informal meetings; observed 16 network meetings, interactions and activities; and analysed over 250 documents. The interviewees were selected through targeted and snowball sampling directed at stakeholders relevant in and for the network. The results of the evaluation underpinning this paper, as well as initial drafts of the papers were discussed with key figures in the network.

All of the data collected (observational field notes, interview transcripts and text-based documents) were coded in NVivo qualitative data analysis software, enabling an integrated analysis. To unravel the relationship between ownership and effectiveness, we focused on collective identity, accountability processes and the advocacy message, as the tensions we found were expressed clearly and distinctly in these three domains.

2.4.1. Case Study

Since its establishment in 2005, GPPAC's main strategy has been networking and enabling civil society worldwide to create a shift from reaction to the prevention of conflict: 'We believe that if we, as a civil society network, influence intergovernmental organisations, state actors, civil society and other conflict stakeholders to create, improve and implement conflict prevention mechanisms in ways that are locally grounded and inclusive, conflicts are more likely to be resolved before violence escalates'.[18] During meetings, we observed the value of ownership to be central, because representatives are given space to engage and share their knowledge with likeminded people from around the world. Members expressed that belonging to a member-owned network feels empowering: 'We as [a network] are not marginal in our own societies and media; we are a voice that can only be heard in cooperation'.[19] In line with Keck and Sikkink (1998), we found that these principled collective values and beliefs around

[18] Interview with network director, 2013.
[19] Interview with network member, 2014.

ownership and change processes are the motivation for belonging to the network. This is part of what identifies GPPAC as a TAN.

GPPAC is made up of a diverse range of regional networks and the individual members of approximately 180 organisations, among whom active membership and levels of involvement vary. GPPAC values shared governance among its diverse membership. This membership ranges from volunteer organisations to research institutes to peacebuilding activist groups, facilitated by GPPAC's regional and global secretariats. The regional secretariats, besides hosting their members, also develop their own agendas, whereas the global secretariat, comprising 16 permanent staff members, must facilitate the whole network and develop a global and collective agenda.

The agendas and priorities from GPPAC's 15 regional networks are the basis for the collective global objectives: 1) to capacitate and enable members to improve, initiate and implement their practice; establish a common agenda and raise awareness on multiple levels; and 2) to capacitate and enhance interaction and to develop good practices and effective coordination with targets (United Nations, regional intergovernmental organisations and state actors). These objectives are meant to 'pool civil society efforts and develop strategic partnerships around these themes with the three key targets' (GPPAC Internal Strategic Plan). Central to these objectives is the achievement of shared ownership as an important vehicle for the anticipated change through four key strategies: network strengthening, action learning, public outreach, and policy and advocacy. The intent of these strategies is to mobilise and link regional, cross-regional and global exchanges focused around the thematic priorities of preventive action, dialogue and mediation, peace education, human security and gender.

2.5. Findings

As a first stage in our research, we studied the network's strategic documents, outlining who they were and what they wanted to achieve. The aim of achieving a shift from reaction to the prevention of conflict by being member-led stood out. However, during the first interviews, it was difficult to grasp how this was operationalised in practice and what it meant to be a member-led network in concrete terms. During our observations of the network's meetings, the network processes became clear. Network meetings were not only moments of network governance; they were also places of lively interactions of diverse but likeminded people who felt inspired and empowered through their membership in the network. A main finding from

these observations was that the network always prioritised inclusiveness over its other objectives.

As we proceeded, we learnt that our difficulties in grasping the nature of the network came from the discrepancy between what the collective, inclusive network pursued and how being an inclusive network worked out in practice. Increasingly, it became clear that the network was both a means and a result. To retain its inclusive properties, regional processes became the dominant dynamic of the network and were far more important than the collective realisation of the network's global objectives. This shaped, to a large extent, the collective identity, accountability and advocacy of the network and influenced what the network could achieve.

2.5.1. 'Being an inclusive network'

> We are a global, member-led network of civil society organisations who actively work on conflict prevention and peacebuilding (Network member, 2012).

What did it mean to be a 'global member-led network' for the collective identity of GPPAC's stakeholders? The members, although diverse in background—ranging from full-blown peace activists to research institutions—and active in varied geographic and thematic areas, felt connected through their shared belief in being an inclusive network: 'Feeling part of the network, ownership and responsibility were important' (regional representative, 2014). Hence, while the subject-matter of the network concerned peacebuilding, the inclusive nature of the network strongly determined its identity as constructed by those involved. According to many representatives in meetings and interviews, being inclusive was the hallmark of their collective identity. This defined who they were, what they did and how they presented themselves; it was thus central in network language, structure and representation.

Members and network staff used the same jargon around what they were (an inclusive network) and what they did (connect, link and pursue a shift from reaction to prevention). Although this seemed plausible (in a be-the-change-you-want-to-see way), we noted that the general language was not easily translated into concrete terms, actions and realisations. Members often chose their wording from strategic documents without unpacking what this meant for their everyday regional realities.

The emphasis on inclusiveness also affected the network's structure and interactions, because it created a certain role division. The idea that the network was member-owned meant that the secretariats were facilitators rather than drivers. Additionally, the global secretariat had

the role of facilitating the network as a whole. Staff members repeatedly stressed that members were 'the drivers' who 'have all the knowledge', and emphasised, 'We do not decide what happens; our members do' (programme manager, 2013). This resulted in a role division that limited what the secretariat staff could demand from members: 'We do not want to be pushy as global secretariat, but [rather] facilitate interaction' (programme manager, 2013). In meetings, we saw how staff members were careful to ask for input, feedback, engagement or reports, thus providing space for regional processes and priorities and conveying a culture of prioritising members. As a result, internal dynamics tended to make the network the sum of its parts rather than a collective.

Aside from driving internal dynamics, the identity of being inclusive provided an important image the network used to present and distinguish itself from others. In the global arena, we observed how the network created access and space for regional voices by advocating for inclusiveness. This opened doors for local peace advocates. In meetings with stakeholders from around the world, network representatives continuously stressed their inclusiveness as a selling-point. Policy officers interviewed in 2014 commended the network's inclusiveness as 'a potential' and 'an interesting asset', stressing, 'You guys [referring GPPAC] have the evidence, and evidence from the ground is important', and 'Engaging with the network is great, because policy and practice on the ground are important for us'. Clearly, the inclusiveness approach distinguished the network from other stakeholders, providing legitimacy and credibility and opening space for its members' voices. Local ownership was not only pursued as a vehicle to achieve objectives;; it was embedded in the entire idea of what the network represented.

2.5.2. Accountability as an inclusive network

> Everything has to be within GPPAC structures, being accountable and representative. Now we are coming to a point where we see, well, maybe that is limiting (programme manager, 2012).

Accountability processes evolved around the network structure, and interactions and were shaped by the culture of inclusiveness. In meetings, network activities were discussed and reviewed based on their inclusiveness, questioning, 'Is this GPPAC?' and 'Are we where we want to be?' These questions emphasised and reinforced the idea that ownership was central, and we observed how decisions on strategies and actions translated this into network practice. Member ownership acted as the major benchmark against which accountability took shape.

In its shared governance structure and interactions, the network pursued horizontal processes in which members were accountable to each other. Regional members were accountable to each other in the regional steering groups. The regional secretariats were accountable to their members and to the global decision-making bodies. The decision-making bodies, consisting of members, were layered to create adequate input from diverse levels within the network, and the global secretariat was accountable to these same decision-making bodies, the members and the donors.

Interestingly, the hard side of accountability (i.e. resource allocation and financial accountability) was separated from these decision-making procedures and bodies. Accounting for funding happened at three levels: Members were accountable to the regional secretariats, regional secretariats reported to the global secretariat and the global secretariat reported to the donor. There was no direct interaction between the donor and the members, and the global secretariat aimed to act as a shield. This created the policy space for members to follow their own priorities, because they were not constrained by donor demands. Consequently, this also meant that formal accountability to the donor occurred in the opposite direction of internal accountability.

This resulted in a culture of inclusiveness, where accountability stemmed from the key value of ownership. Inclusiveness being the dominant construct of the network influenced accountability interactions. The facilitating secretariat did 'not want to be pushy', and members were 'not to be overburdened with demands or questions' (from interviews with programme managers, 2013; and meetings with evaluation manager 2012, 2013). Another programme manager stated that '[a]ccountability is really an issue and sensitive, because we do not want to be top-down and we do not want to be too critical'. Members' freedom to manoeuvre was inhibited only by the monitoring and evaluation system installed to enable accountability to the donor. This system was perceived as a burden: 'There are problems with the amount of demands on members. The monitoring and evaluation system is considered extensive and too much. Also regional members are not interested in reporting. This provides problems with the quality and information flow' (programme manager, 2013). Consequently, ownership-based accountability inside the network challenged the requirements for formal accountability to the donors.

In this way, accountability based on information flowing throughout the network was both shaped and challenged by the network's inclusiveness. Regional and programme reports often provided rich information on activities, but this varied greatly between regions. Additionally, these reported mostly on internal network activities such as annual meetings organised or agendas developed. Although the network pursued this kind of reporting to promote

inclusiveness, the structure did not leave much room for reflections on activities, collective objectives or the diverse roles of actors in the network. Moreover, there were no consequences based on this reporting structure. Therefore, reporting was mostly a bureaucratic exercise with limited references to GPPAC's collective objectives and strategies.

This was illustrated by our fieldwork in one of the regions and at the global level. At an annual regional steering group meeting in Southeast Asia, although we noticed the significance of member interactions as a form of learning and sharing, we also saw that the meeting was not used to ask members to account for the agreements made as a network or to define the achievements in view of the broader objective of preventing conflict. The global collective agenda only figured in the meeting's introduction; progress with this agenda was not actually discussed, as no one probed the members about their contribution to the collective goal. This made it difficult to pursue accountability on commitments made, because information about these commitments and achievements was not provided and no questions were asked on these matters. Similarly, at global-level meetings, members were free to decide what to report, and there was no assessment related to a results framework to which members were held accountable. Hence, the meetings were about being a network more than discussing results.

2.5.3. The inclusive advocacy message

> We do not have an advocacy message; the network is the message (programme manager, 2013).

This statement illustrates that ownership became the network's message, because pursuing inclusiveness demanded a translation of the question: 'Whose message is being promoted?' Network members and staff mentioned there was no specific thematic content-driven advocacy message, because the message was the network itself and being inclusive. This created tension between the network's objective to 'create space for civil society' at policy levels by 'being a member-owned network' and the pursuit of a collective advocacy agenda for conflict prevention. This tension was discussed in internal global secretariat meetings. For example, a programme manager questioned, 'Does it need to be relevant to everyone all the time, or should we be focusing […]? That is the big existential question for a global network'. In other words, a recurring question for the network was 'Are we spreading [ourselves] too thin?' However, although the global secretariat reflected on these questions in internal communications, members were not confronted with these issues.

Being an inclusive network was the main path to creating space for civil society. As a programme manager explained, 'if we contribute to establish[ing] access for civil society to the structures of regional and global policy arenas, it is an important result'. This meant that decisions made to ensure inclusiveness resulted in a wide range of advocacy:

> GPPAC wasn't set up to achieve a specific goal, such as, for example, eliminating landmines, which would need a coordinated targeted campaign, but GPPAC is as much about building capacity of members and how they do things around the world as making an impact at UN Headquarters. We are talking about very diverse measures of impact, and sometimes this is spread out (programme manager, 2014).

Rather than coordinated advocacy strategies to influence policy and practice, network members considered being part of GPPAC as providing 'international-level voice' and 'opening doors' (members from West Africa); setting out a 'common platform useful for connecting people' (member from the Middle East); and offering 'clout, prestige and input' (member from Southeast Asia). By facilitating members through 'an ongoing structure that hopefully increases the capacity of the members in every way' (programme manager), the network pursued embedding its advocacy in local knowledge and content and advancing these regional efforts by providing worldwide connections to build legitimacy, outreach, space and capacity (from interviews with global members, 2013; and programme staff, 2014).

In reality, we observed that this resulted in parallel processes. Member capacities and thematic priority areas varied greatly at regional level. Whereas one region focused on peace education, another was more concerned with bringing conflicting groups together for dialogue and mediation. At the global level, the network pursued collective objectives by linking regional voices to policy arenas, but the international activities were limited in scope, so only a few issues from the regional agendas could be brought to international arenas. Members' engagement selecting these was limited, and the global secretariat's agenda was set mainly by opportunities, internal considerations of the secretariat staff and a few particularly active and influential members.

This was observed at global level, where GPPAC annually enabled some members to strategize and advocate around the UN gender week in New York. As these members all had their own regional priorities and because the meetings were member-driven, developing a collective agenda took a significant amount of time. Most members were generally dependent

on the global secretariat to set things up and move forward. Eventually, a few members, who had their own global agendas aside from their involvement in the network, took the lead. Therefore, the value of the network in New York consisted primarily of opening doors; regardless of the specific agenda a member wanted to address, the global secretariat facilitated meetings using the argument of ownership to garner legitimacy. However, despite opening doors, translating the regional priorities into direct policy and practice recommendations remained limited. These meeting were singular instances of influencing, with limited results-oriented follow-up.

In addition to facilitating networking during international meetings, a number of thematic programmes were developed to pursue cross-regional engagement and collective advocacy. However, global-level collective processes were not prioritised by network members. Although some interviewed regional members saw opportunities provided by the global network, others stressed that the priority was their regional or local work. Members found the global-level processes to be disconnected from their everyday realities and the global objectives too abstract. Moreover, cross-regional engagement was often limited to a few members and did not necessarily trickle down into the regions or up into the global collective. Hence, the global-level advocacy was done mostly by the global secretariat. Because the secretariat could not speak for the members or form a message on their behalf, it tended to frame its mandate in terms of building relations, convening and learning.

2.6. Being inclusive and effective

> Members need to relate to GPPAC and it is not important to relate to themes and strategies but rather to relate to its bottom-up, unique approach and the basics. There is no particular agenda on how to do things. That is not what we do and not important to them; it is about ownership (regional representative, 2013).

As was illustrated above, the network's emphasis on ownership shaped its collective identity, accountability and advocacy, which influenced the basis of the network's effectiveness. The assumption was that the member-owned nature of the network meant that effective advocacy messages would evolve from its inclusiveness. All of the interviewed members underlined the importance of belonging to the member-owned network. Representatives commonly mentioned the importance of linking and learning through the network, arguing that GPPAC opened doors

and provided access to other civil society organisations and potential targets, as well as offering spaces for interaction.

In Southeast Asia, we observed how the annual regional meeting facilitated members to interact on diverse subjects, share their local perspectives and learn from each other. The regional secretariat hosted an outreach trip around this meeting, bringing together members, external partners and policy stakeholders to discuss a specific local conflict. This provided an opportunity for members to see how other stakeholders dealt with conflict and to connect with regional policy makers and local peace activists. This inspired learning, as participants found connections and similarities to their own situations. These annual meetings also inspired members to form local networks—for instance in Burma and Southern Thailand. As elaborated above, most of the results from these meetings were internal networking and exchanges, and it remained unclear what tangible results on conflict prevention could be claimed, reported and proven.

Although inclusiveness inspired members, created access and space, and brought legitimacy and credibility to the network, it also challenged the effectiveness of the network's collective advocacy. Shared ambitions of achieving 'a shift from reaction to prevention of conflict' were translated into broad and thematic collective goals on the global and regional levels without defining desired results that could be planned and monitored. It was difficult for members and staff to identify actual results that the network had achieved, as opposed to what they did as individual members. When asked for results or evidence of achievements in relation to the collective ambitions, representatives tended to repeat jargon from collective documents or stress the importance of being a network in abstract terms.

These observations call into question the widely shared assumption in GPPAC that advocacy content would follow from the inclusiveness of the network. Although there were a number of cases where the network did develop content—on the issues of gender, human security and the post-2015 agenda—this remained relatively broad and abstract to prevent precluding any issue that might arise from the members and regions. As a result, for members, the collective goals were often rather 'disconnected from everyday reality in the regions' (network members during an internal network meeting, 2013). The broad diversity of the thematic issues and geographic scope, along with the absence of any prioritisation between them, created challenges related to 'spreading too thin'. This constrained the articulation of a clear message. A programme manager stressed that 'our minimum is being a likeminded organisation'; the '[a]dvocacy is also not focused but is directed at processes and being part of the processes with our targets, silent diplomacy and setting the agenda together'. Regional successes in the network could

largely be traced back to the efforts and capacity of specific members, rather than the network. In some cases, interactions at regional level were meaningful; however, the contribution of the network mostly comprised financial support and organisational efforts by the global secretariat.

The level of inclusiveness caught the network in a bind. On the one hand, the secretariat could not pressure members to provide information or commitments. On the other, the secretariat was also unable to push advocacy issues without its members actively engaging, because their role was purely facilitation; 'The knowledge is with [the members]', 'this is [the members'] process' and 'you [the members] are its managers'. By placing members at the centre, the network prioritised creating space for meaningful connections at regional level and between regional and global processes. This included space for partnerships and connection, learning and sharing amongst likeminded members from around the world. However, there were no requirements regarding members' commitment to the network, and there were few incentives for this commitment. As a result, 'belonging to the network is valuable, but engaging in activities is not' (programme manager, 2014). The network secretariat thus struggled to ensure commitment, as the expectation that commitment and message would automatically follow ownership did not materialise. The gap between global ambitions and actual activities created tensions within the network that were discussed within the global secretariat but were rarely mentioned in broader network meetings.

The network's global ambitions were also challenged by the diverse levels and spaces where it operated. While the global ambitions illustrated the pursuit of being a collective network, in reality, the network was often a 'partial collective'(network director, 2014). For example, work on regional issues, such as the North Korea peacebuilding mission or the conflict in the South China Sea, does not necessarily need to involve the network as a collective, but rather the parties directly impacted. Tensions around commitment and partial collectiveness resulting from the ownership dynamics affected GPPAC's relationship with donors and other stakeholders focusing on rational, tangible and results-based effectiveness within accountability frameworks. The network approach, while broadly appealing in terms of ownership, did not fit such frameworks.

Overall, the examined network was highly visible and capable of achieving network-related influence (i.e. networking, linking, convening, relations building, knowledge sharing, access and space for regional voices), but this was not matched by observable impact in content-based advocacy beyond what members were achieving on their own (i.e. campaigning, collective claims, content-driven advocacy). Moreover, the strong culture of inclusiveness made it impossible to reflect on this tension during network meetings. Although individual members

and the secretariat staff expressed concerns about this in interviews, they repeatedly stressed that the network's role was merely facilitative. They appeared to see no room for creating space in regional- or global-level meetings for discussing what this meant for the network's ambitions and practice in relation to its mission of conflict prevention.

2.7. Conclusions

In the development field, shared ownership is considered a condition for effectiveness. This has the characteristics of a mantra, as it is a belief underpinning policy and practice without being made explicit or being put to the test. The consequences of the political, constructed nature of ownership and effectiveness and the inherent contradiction between them is rarely discussed. The contrast between effectiveness, as an instrumental concept treating results as unambiguous and rational, and ownership, as a social construct acknowledging the importance of different perspectives, makes understanding how the two concepts operate together in practice quite complex.

This study calls for a more careful consideration of the assumption that ownership results in effectiveness. In the case of GPPAC, we observed real tensions between the two concepts, with implications for the network's effectiveness. In terms of networking, GPPAC was highly effective, but, in terms of the mission of conflict prevention, the effectiveness was much more diverse. These differing notions of effectiveness often played a role in the undercurrent of network meetings, but they were not explicitly discussed, as open reflection about the meaning of effectiveness was curtailed by the strong ownership narratives and norms. There was little room to raise discussions that could appear to undermine or question the core value of ownership.

Several studies have discussed links between organisational structure, network management, power and effectiveness. These studies have demonstrated that balancing and managing diversity in relation to the collective is crucial for networks to be effective in achieving their advocacy objectives. The present study extends this discussion by further unpacking the relationship between a stated objective and the network's everyday practice, which establishes what effectiveness actually means to the members of the network. By focusing on the processes of network identity, accountability and advocacy, we reconstructed the interplay between effectiveness and ownership.

Our study partly confirms the findings of Saz-Carranza and Wong indicating that effectiveness requires a balance between network unity and diversity, as both provide the

network with the power of legitimacy, credibility and the ability to take action: Inclusiveness nurtured GPPAC's diversity at the expense of seeking collective objectives that would require more coherence and unity. However, whereas the nature of effectiveness was left undisputed in Saz-Carranza's and Wong's work, our findings revealed that a strong emphasis on ownership may generate a different understanding of effectiveness, where the formation, operation and linkages of a network can be seen as both the means and the result of advocacy work. Consequently, the network generated value by providing space to local realities and everyday practices and by creating and sustaining relations with targets and amongst members where mutual learning and sharing thrived and feelings of belonging to a network of likeminded partners empowered and inspired members.

The main lesson learned from this case study is that, although ownership and effectiveness are related, in practice, these concepts are ambiguous, while continually challenging and defining one another. Effectiveness is a multidimensional concept that acquires meaning through interactions among social and political considerations, everyday practice in the operational environment, and implicit values and norms. International advocacy networks should be aware of the ambiguous nature of effectiveness as it relates to ownership. It is important for networks to make these tensions and dilemmas explicit points of discussion and to integrate their understandings of these concepts and how they relate to their ideas about change. This is especially relevant today, as networks are becoming increasingly prominent in the development agenda. It is important for future work to further unpack the tensions related to these collaborations, such as that between resource distribution and ownership.

For international advocacy evaluators and researchers, our findings imply that set definitions of ownership and effectiveness cannot be taken for granted, and an automatic and positive association between the two concepts cannot be assumed. Rather, it is necessary to investigate how networks define the concepts and how the relations between them play out in practice.

3. Negotiating effectiveness in transnational advocacy evaluation[20]

International development organisations increasingly use advocacy as a strategy to pursue effectiveness. However, establishing the effectiveness of advocacy is problematic and dependent on the interpretations of the stakeholders involved, as well as the interactions between them. This chapter challenges the idea of objective and rational evaluation, showing that advocacy evaluation is an inherently political process in which space for interactions around methods, processes and results defines how effectiveness is interpreted, measured and presented. In addition, this chapter demonstrates how this space for interaction contributes to the quality and accuracy of evaluating advocacy effectiveness by providing room to explore and address the multiplicities of meaning around identifying, measuring and presenting outcomes.

Keywords: Advocacy evaluation; negotiation; outcomes; effectiveness; international development

[20] This chapter was accepted as article with minor revisions and published by Evaluation as: Arensman, B., and Wessel, M. V. (2017). Negotiating effectiveness in transnational advocacy evaluation. *Evaluation*, 1356389017733210

3.1. Introduction

In the last two decades, nongovernmental organisations (NGOs) have increasingly used advocacy as a strategy for effectiveness, pursuing social and political changes. Such processes and their results, however, are not easily measured or predicted (Devlin-Foltz et al., 2012). Advocacy evaluation is a nascent field that challenges conventional evaluation methods, which often seek to measure and predict results in quantifiable terms. NGOs are under strong pressure to present these quantitative results through evaluation,1 which is assumed to be objective and independent, to demonstrate legitimacy and account for funding streams (Eyben et al., 2015; Riddell, 2014).2 A deeper understanding of advocacy evaluation is needed, because evaluating these efforts and their influence demands different evaluation approaches.

This chapter makes three central arguments. First, advocacy is extremely difficult to evaluate. In this context, the relation between funds and achievements is ambiguous (Roche and Kelly, 2012a, 2012b). Advocacy is about influencing social and political processes, behaviours, policies and practices (Keck and Sikkink 1998). This often happens in informal settings, with few opportunities for establishing and tracing evidence. Advocacy demands acting and reacting quickly in response to opportunities and threats. Therefore, advocacy processes cannot be planned in a straightforward manner, and advocacy cannot be evaluated by measuring processes and change against a preconceived set of planned deliverables.

Second, advocacy evaluation is necessarily subjective. Drawing on our experiences evaluating eight transnational advocacy programmes, we show how effectiveness was necessarily identified through direct interaction and negotiations with staff members of the evaluated organisations. These organisations were highly interested in favourable results, so establishing effectiveness was of major concern. This chapter discusses how this challenge manifests in practice and identifies important implications for advocacy evaluation.

Third, the subjectivity in evaluating advocacy provides interesting opportunities to probe, reflect and develop insight into the value of advocacy work. Unavoidably, interested parties influence the evaluation through interaction, but this can contribute to evaluation quality and accuracy in important ways. We argue that, rather than an exercise in objective assessment, evaluating advocacy is a dynamic and political process that takes shape through interactions. Our findings contribute to discussions on the tension between the pressure for measurable results from objective and rational evaluation versus evaluation as socially and politically constructed (Eyben et al., 2015; Riddell, 2014; Taylor and Balloch, 2005).

We used a large evaluation of transnational advocacy programmes as a case study to unpack how effectiveness is given meaning through interactions. We focused on several key questions: *How do interactions in advocacy evaluation give meaning to advocacy effectiveness? What does this mean for the identification and measurement of outcomes (i.e. access to evidence)? To what extent do negotiations, dialogue and co-creation shape these processes and outcomes? How do these negotiations affect the presentation of outcomes (i.e. negative outcomes or positive framing)?* To answer these questions, we focused on patterns of interactions around three important aspects of evaluation: the identification, measurement and presentation of outcomes.

We begin by discussing key debates on evaluating effectiveness, realist positivist/post-positivist versus constructionist traditions, and advocacy evaluation as socially constructed. Next, we describe the methods used and introduce the case study evaluation. We then present our findings on how effectiveness in advocacy evaluation is shaped by interactions about identifying, measuring and presenting outcomes. The conclusion draws out broader implications for theory and practice on advocacy evaluation and effectiveness, as well as directions for future research.

3.2. Advocacy challenges realist evaluation

3.2.1. The results-oriented approach

Growing pressure for results has pushed NGOs and donors into evaluations based on traditions of prediction and control, seeking general patterns of cause–effect relations (Armytage, 2011; Riddell, 2014), based on the assumption that evaluation is value-free (Marchal et al., 2012). Increasing challenges to the value and effectiveness of development have resulted in new forms of management aiming to improve effectiveness based on performance and results-based regulatory frameworks, including evaluations (Guljarani, 2011). These developments reflect dependencies, expectations and a results paradigm, all of which falsely imply an ease of objectively evaluating and delivering outcomes. This approach demands replicable short-term outcomes aligned with the intended objectives stipulated in funding agreements (Guljarani, 2011; Riddell, 2014) and the development of accountability-driven tools and methods that work with predefined indicators to demonstrate effectiveness through measurable and tangible results (Alexander, Brudney and Yang 2010; Eyben et al., 2015).3

These demands inspired theory-based and objectivity-based evaluation (Lam, 2002) from a rational–realist approach.

Realist evaluation is part of the post-positivist tradition. Whereas positivism assumes an independent relationship between the evaluator and the evaluated, post-positivism pursues objectivity but accepts that contexts and mechanisms influence outcomes (Mansoor, 2003; Van der Knaap, 2004). Realist evaluation goes beyond merely focusing on the success or failure of interventions; it acknowledges that events are rarely the result of a single causal mechanism and that outcomes are not independent of context (outcomes are context + mechanism) (Pawson and Tilly, 1997; Sayer, 2000). The realist approach systematically tracks causal relations between mechanisms and outcomes, seeking explanation and control. This implies that a causal relation can be objectively defined and established. This theory-based tradition does not consider the role of the evaluator or of the interactions between the evaluator and the evaluated in shaping the evaluation outcomes. The tradition sees evaluation and evidence as neutral and rational, failing to acknowledge social and political construction.

In his criticism of realist evaluation, Porter (2015) has noted that this approach to outcomes ignores agency and behaviours of interest (For whom are which outcomes desirable or undesirable? What are the effects of the interventions on those affected by them? What are the effects of the evaluation outcomes for those involved?). The positivist tradition in evaluation is characterised by a search for the truth. However, not everything can be measured in the hard sense of the word. Constructionists emphasise that this perspective is too simplistic and does not do justice to the complex and dynamic nature of reality. They argue there can be no generalisable or universal claim of truth or reality (Van der Knaap, 2004).

Taking a realist approach to evaluation risks overlooking how effectiveness is formed and understood, as well as the role of underlying dynamics such as relations, interests and interactions in shaping this process. Such dynamics are also the politics of evaluation. Evaluation is political in terms of decisions about which outcomes are included and which define effectiveness, and about what is measured and how. Regulatory frameworks are defined by those in power, who demand results from those who must demonstrate these results for organisational and professional survival (Eyben, 2015). Approaching evaluation from an objective perspective does not allow for challenges to these power relations, as it assumes evaluation roles, concepts, purposes, methods and outcomes are non-negotiable. According to Riddell (2014: 4–6), this hampers a thorough understanding of effectiveness, because it does not provide enough information to draw meaningful conclusions about what is actually achieved, what works, how and why. As a result, underlying problems with evidence, reliability

and validity, the collection of adequate data, the demonstration of attribution or contribution, the establishment of counterfactuals, and access to information are not addressed.

3.2.2. Advocacy evaluation as socially negotiated

In evaluation studies, a growing body of literature challenges the results-focused tradition and seeks an understanding of underlying dynamics and complex processes, such as advocacy, which are not easily captured by positivist designs. Guba and Lincoln (1989) provided an early demonstration that there are many diverse perspectives that evaluators need to consider. Research from a constructionist perspective emphasises that all evidence (i.e. knowledge) is contextual, relative, and subject to interpretation and power, challenging the notion of objective, independent and rational evaluation (Eyben et al., 2015; Hayman et al., 2016; Taylor and Balloch, 2005). This understanding demands critical reflection on evaluation theory and practice, also because of growing concerns regarding NGO legitimacy and effectiveness, which evaluation is key in assessing. Ringsing and Leeuwis (2008) have argued that, to improve how monitoring and evaluation are set up and pursued to create space for reflexivity and learning, both affirmative political action and leadership are required.

An evaluator is more than a gatherer and assessor of data. An evaluator is also a negotiator and facilitator, managing and navigating learning, expectations and interests (Markievicz, 2005; McDonald, 2008; Sharkey and Sharples, 2008; Taylor and Balloch, 2005). In advocacy evaluation, such an approach is highly applicable, but existing theoretical and empirical work has not explored these roles in the evaluation of advocacy.

Advocacy processes are often long-term and transnational, involving multiple stakeholders. In development processes, advocacy mostly concerns complex transformational changes. In this context, causal relations between actions and results are difficult to establish, and changes achieved tend to be largely invisible and hard to trace. Influencing often takes place behind closed doors, and those influenced by advocacy may not always be willing or available to disclose whether they were influenced by specific actors, actions or events. Moreover, intervention effects are found among numerous additional causal strands. The objects of advocacy – policy makers, publics and private sector actors – are moving targets and are subject to numerous influences. Hence, establishing links between a change and an advocacy programme is often difficult (Roche and Kelly, 2012a, 2012b).

The present research resonates with previous work emphasising the negotiating and facilitating roles of the evaluator (Markievicz, 2005; McDonald 2008), the political character

of evaluation (Roche and Kelly, 2012a, 2012b; Taylor and Balloch, 2005), and the necessity of a collaborative, participatory, negotiated and facilitative approach to evaluation (Liket et al., 2014; Patton, 2011). These studies have stressed diverse elements of interaction as important in evaluation. Taken together, the findings of existing work on evaluation in general and on advocacy evaluation highlight the need to clarify the meaning of assessing effectiveness in the evaluation process in terms of interactions driven by relations, interests and power. Accepting the interactions involved, we incorporate negotiation as inherent to the evaluation process.

3.3. Methods and case

The evaluation used as a case study in this chapter was carried out from 2012 to 2015. This evaluation was commissioned by the Netherlands Ministry of Foreign Affairs (NL MFA), in cooperation with the Foundation for Joint Evaluations (SGE), which represented the evaluated Dutch development organisations. The evaluation was coordinated by the Netherlands Organisation for Scientific Research and included eight transnational advocacy programmes receiving NL MFA funding. The evaluation team consisted of 12 evaluators with diverse profiles, including academics, evaluation and thematic experts, and consultants with multiple years of experience with international development processes.

The main purposes of the evaluation were to account for the results of the included programmes, to contribute to improving future development interventions and to develop advocacy evaluation methodology. The evaluation team was tasked with addressing predefined evaluation questions on the changes achieved, the contributions to these changes, the relevance of the changes, the efficiency of the efforts made and the factors explaining the findings. The team was instructed to focus on three predefined priority result areas: agenda setting, policy influencing and changing practice, mirroring the expected (but not necessarily linear) phases to reach the advocacy programmes' objectives on changing agendas, policies and practices.

The analysis presented in this chapter is based on the authors' reflections on the evaluation process; it is not the analysis of the evaluation team. Although the team was closely involved in learning and developing our approach, there was limited time for more fundamental reflection on advocacy evaluation by the full team. The issues addressed in this chapter arose in discussions between the authors during reflections on the processes in the final stages of the evaluation and afterwards. The evaluation assignment included learning as important aspect, and the evaluation was part of the first author's PhD research, allowing additional space for constructive analysis.

The evaluation drew upon document review, semi-structured interviews and (in-) formal meetings with representatives from the evaluated NGOs, as well as with other stakeholders such as their partners, advocacy targets (policy makers, private sector organisations and civil society) and other relevant stakeholders in the field (Arensman et al., 2015). Where possible, participant observation was also conducted.

In this chapter, we focus mainly on the three advocacy programmes the authors were closely involved in evaluating. We followed the diverse advocacy programmes through time (2012–2015) and space (diverse geographic locations and political arenas worldwide), conducting a multi-sited ethnographic field study (see: Marcus, 1995). The analysis for this chapter is based on 300 interviews and formal and informal meetings with evaluated programme staff, partners and advocacy targets, as well as the observation of 29 meetings relevant to the advocacy programmes. To embed the three cases in the wider evaluation context, we also drew on three meetings relevant to the evaluation process that included all of the diverse stakeholders involved, as well as notes and reflections from 15 internal evaluation team meetings.

For the analysis, the data were systematically searched for phenomena and socio-lingual aspects providing meaning to how advocacy was pursued and how outcomes were perceived. This was an iterative process of clarifying the meaning of interactions, words and perceptions in interviews and exchanges between the evaluators and the evaluated, and among the evaluation team. This approach to the analysis allowed space for multiplicities of meaning regarding the identification, measurement and presentation of outcomes. The analysis was assisted by NVivo qualitative data analysis software.

In line with our argument that there is no such thing as an objective evaluation of advocacy, we understand that our position as evaluators and academic researchers was not value-free. As evaluators, we were insiders in the evaluation process, able to access privileged information on decision making and negotiations. In our capacity as evaluators/researchers, we were also observers reflecting on the processes. As external evaluators, we were outsiders to the evaluated programmes, and this influenced what information we could access and how we could access it. The next section presents our analysis and findings uncovered as evaluators and researchers navigating interests, access, co-creation and learning.

3.3.1. Creating space for interaction

The evaluation framework created space for interactions that shaped the evaluation in a co-constructed manner. The evaluation team worked with a predefined framework, but the

framework did not specify in detail exactly what was to be evaluated or how this should be done, because the evaluation team was expected to make a methodological contribution to the nascent field of advocacy evaluation.

In the many discussions amongst the stakeholders involved during the first year, fundamental questions arose: What constituted advocacy outcomes? What kind of information was appropriate and available to assess outcome claims, and what level of evidence was necessary and reliable? There were no self-evident answers to these questions. The open-endedness of the framework in terms of conceptualisation and methods allowed for progressive development and adjustments, contributing to learning about advocacy evaluation.

As evaluators, we were challenged to pursue independence, while seeking accountability and learning and navigating access to information. In this balancing act, many interests were at play, with stakeholders' legitimacy and credibility clearly at stake. For the funder of the advocacy programmes (the NL MFA), a timely and rigorous evaluation that objectively established effectiveness was very important. The evaluated programmes' main interest was in meeting the donor's accountability demands to legitimate the programmes and their funding and to show the effectiveness of their work, encouraging possible future funding opportunities. For us as evaluators, key interests were producing a quality report, staying within budget, doing justice to the NL MFA's requirements and to the advocacy programmes, and learning to develop advocacy evaluation approaches and methods. These interests, along with the open-ended framework, provided space to develop the evaluation methodology, process and results through interaction.

The evaluation framework required the use of specific result areas in answering the predefined evaluation questions, which centred on results. However, the development of our approach and methods demanded openness and flexibility in our engagement with the programmes. To gather information and comprehend the advocacy processes, we needed the support of the programmes. Programme staff sought engagement with us to develop the evaluation to fit their needs. In interaction with the evaluated programmes – partly of our own accord, but frequently initiated by the SGE and the evaluated programmes' staff – we pursued dialogue, seeking to include diverse perspectives on the methods, process and results. In 2012, we organised a stakeholder meeting to share and adjust our ideas by gathering input via open discussions, assuming the stakeholders would provide valuable insights and knowledge on how to deal with advocacy evaluation. Moments to share and learn also happened in formal interviews and informal and formal meetings. These occurred at various levels: within the

evaluation team, between the evaluated programmes' staff and the evaluators, and between the coordinating agencies and the evaluators.

In the following section, we illustrate how this space created for interaction among the stakeholders involved in the evaluation gave meaning to our understanding and assessment of effectiveness. First, we explain how interactions about *identifying* outcomes shaped our understanding of outcomes and our ability to identify them. Second, we show how ways of *measuring* outcomes were constructed through negotiation. Third, we explain how the interactions about our *presentation* of the outcomes influenced the evaluation findings in the final report.

3.4. Shaping the evaluation process and outcomes

3.4.1. Identifying outcomes

Conceptually and practically, organisation staff and evaluators interpreted outcomes in different ways. Our interactions provided space for dialogue and negotiation regarding these interpretations. These exchanges were a crucial part of the evaluators' efforts to establish what the programmes had achieved. Here, we present two examples of interaction processes, offering insight into the broader patterns of co-constructing, discussing and negotiating the identification and meaning of outcomes in the evaluation. Overall, these interactions extended our understanding of advocacy processes to include outcomes that were not necessarily claimed or visible, as well as networked outcomes.

Co-created outcome identification

The evaluation team defined outcomes as follows:

> […] changes – intended or unintended – in the three priority areas of agendas, policies, and practices, as well as in networks and relationships, in governance structures and processes, or commitment and involvement of a particular actor. These changes must be observable and traceable to the advocacy and lobby activities under review and may also include negative changes that happened. (Arensman et al., 2013: 20)

During the evaluation, we learned the concept of 'outcome' was subject to multiple interpretations, resulting in dialogue about the nature of outcomes and processes of co-creation to identify outcomes. One advocacy programme based in the global South provided us with a thick booklet presenting their main achievements. These were mostly activities they organised, key stakeholders who joined their board and specific outputs (i.e. research reports and policy briefs). The booklet also included one example of an African government that improved its child rights policies after they were 'named and shamed' in a ranking study by the organisation. Although this provided meaningful insight into the wide diversity of projects the organisation conducted and their understanding of achievements, it did not include explicit changes to which their advocacy had contributed. We learned that they understood outcomes to be within their direct sphere of control. As evaluators, we were looking at programmes' 'sphere of influence'. We therefore had to reconstruct the processes around their advocacy to surface the outcomes achieved. This included exploring their interpretations and explanations of what they did and why, as well as how it was strategized, in addition to our questions about outcomes as changes in the sphere of influence.

We recognised that staff members were often unaware of the roles they played in influencing the outreach and uptake of the advocacy message, because they were focused on their direct activities and output. However, through our conversations, we also saw the advocacy work from their perspective; we were able to identify further achievements and understand why the advocates did not initially present these to us as outcomes. For example, we learned that one of their advocated issues was taken up by a specialised committee of the United Nations, and we identified the outcomes together:

> *Interviewer*: What you explain, the invitation by the Committee in Geneva to present the issue you advocate, is already an outcome.
>
> *Interviewee*: Oh… Yes, that is an outcome [surprised]. What has happened, once they [the UN Committee] had agreed, we developed a work plan on what they should be doing. We have monthly teleconferences in the consortium. We have been working very hard on it. What happened, there was a need to develop the working plan, and we developed it and shared it.

Although this organisation had successfully created space to discuss the advocated issue at policy level, and the issue was placed on the agenda of a significant institution, this was not reported anywhere as an outcome. It was only through the interviews that we surfaced what had happened and how the organisation had influenced the process.

Furthermore, we found that what we perceived as an underrepresentation of outcomes was in fact a deliberate strategy to accommodate underlying issues in the working context. First, in the country where this organisation worked, space for civil society was restricted: Even using the term 'advocacy' was problematic. Second, many stakeholders had been involved in advocacy around the issue for a long period, making it difficult to claim a certain influence or change: The organisation was not the only stakeholder involved, and relations/partnerships were politically sensitive. This required taking care not to step on anyone's toes. Moreover, 'non-claiming' characterised this organisation's way of working, and they were applauded for this by the stakeholders and policy makers with whom they worked: '[The organisation] is modest and does not need to claim like some of the other CSOs; this is important because our political representatives are sensitive' (African Union official, 2013). Third, the issues they advocated were politically sensitive. We found that the organisation was able to open space for discussions on these issues precisely because of their high regard for these sensitivities. Thus, the underreporting actually contributed to the advocacy outcomes.

This organisation's reluctance to claim influence or changes was understandable, but it left the organisation unaware of the extent of their successes, and it made their achievements invisible. In discussions with the evaluators, the Dutch partner organisation emphasised that they had no idea whether this programme was effective, and we would never have found the outcomes without taking an open and flexible approach to the narratives of the staff and targets.

Networking as outcome

Our engagement with networking is another example of co-creation in the identification of outcomes. Most of the advocacy programmes worked through, or as, networks. Some were loosely connected networks of like-minded organisations or ad hoc cooperation efforts, whereas others were more permanent organisational structures. Networks were a result in themselves, creating space for cooperation and knowledge sharing. Networking was thus strategically significant to the evaluated advocacy programmes. In the identification of outcomes, this generated questions regarding the role of networking.

Identifying networking as an outcome became an issue of discussion in our team. We questioned how to identify networking outcomes as part of the framework of priority result areas. There was no clear agreement within the evaluation team, and a similar absence of agreement was seen in the conceptualisations and operations of the evaluated programmes. For some, networking was not an outcome, because, in itself, it did not contribute to the programmes' objectives of change. For others, networks were achieved constructs helping to generate entry into policy discussions, and networks could therefore be seen as outcomes. Discussions led to fundamental questions: Was it a matter of the strategic intent behind networking? Or was it important to consider the achieved result of a built network – rather than *networking* – as the outcome?

This was not an academic matter, because, in some cases, these conceptualisations touched upon the core of organisations' work. Networking was a prominent strategy, but it was also actively considered a result in itself. In one of the evaluated programmes, staff claimed recognition for this understanding, presenting us with outcomes such as relation building, network strengthening and knowledge brokering. They argued that the three predefined priority result areas were 'difficult to match with the way networks are functioning' (programme manager, 2013) and risked overlooking the network as the 'backbone for achieving impact [locally]' (advocate, 2013).

Other organisations considered networking a strategy rather than an outcome; to them, it was an integral part of advocacy in a globalising world of increasingly horizontal structures and movements. They considered outcomes to be changes in agendas, policies and practices because of advocacy. Networking was included among a range of strategies. Although networking did not necessarily yield clear or visible outcomes of influence or change, it was considered important for changes in the long run. As evaluators, we had to consider how to deal with the fact that the results of influencing were not always evident, while still doing justice to the significance of networking in the different ways it was presented by the evaluated organisations.

3.4.2. Measuring outcomes

What we measured and how we measured it depended on the evaluation framework as much as the realities we faced. The framework began from the assumption that outcomes are measurable. However, it did not include conceptualisations of what was to be measured against what, or clarifications of how and to what degree proof of achievements was to be established.

In interactions with the evaluated programme staff, we found tensions between the evaluation framework and the realities of intangible, multi-stakeholder and long-term processes in advocacy work. These interactions, together with the limitations in the framework, provided the necessary room for manoeuvre to develop an approach to measurement as the team saw fit. In the following paragraphs, we show that gathering and interpreting information and evidence was negotiated and thus shaped by interactions. As evaluators, we had to accept that negotiations were part of the evaluation process. Key themes that emerged were access to information for establishing effectiveness and the level and nature of 'evidence'.

Access to information

We depended on the willingness and capacity of the evaluated organisations to share information, which meant that we needed to build trust. This was especially important because of the fluidity of advocacy, which happens in quickly evolving environments, on multiple levels and often in informal settings where personal interactions are key. Much of the advocacy work could not be observed through tangible outcomes or followed in real time.

As we were external evaluators contracted by the major donor of the evaluated programmes, the organisations under review were sceptical about our accountability-driven framework. They considered the results to be in tension with their everyday practices. One representative stressed, 'we had these abstract goals, but in reality we do so much more' (2013), emphasising that there were two different realities in which they had to operate. One reality was the everyday practice of advocacy, where advocates had to act and react quickly in response to emerging opportunities and threats. The other reality was that of relations to the donors, which demanded that certain predefined commitments were translated into activities and results.

In our evaluation, we were aware of these multiple realities and demands embedded in the interests at stake, and we sought to relate to these dialogically. We emphasised that we were not a 'fault-finding mission' and attempted to build trust and create a safe space for openness. In this way, we tried to incorporate a better understanding of the complex reality of advocacy and to do justice to the programmes' work.

However, despite these efforts, we were still assessing the advocates' work. Our interest was in truth-seeking regarding the achievements of the organisations. Our report of this assessment could be more or less positive, potentially with significant consequences for the evaluated organisations. Therefore, handling our search for information continued to involve strategic choices by organisations and strategic manoeuvring from our side to meet our information

needs, in the face of conditions that were at odds with the official, donor-imposed requirements of the organisations.

In many cases, the staff of the evaluated programmes did not easily come forward with the necessary information. This was particularly the case for a network organisation whose lead organization (global secretariat) was the direct contract partners of the NL MFA. The members they supported and collaborated with worldwide, who were not directly under contractual obligation, were not in a clear accountability relation with the NL MFA or with the evaluators. This made us dependent on sheer willingness, with some support from the network headquarters, to provide information that could substantiate outcome claims.

In one case, a programme manager stressed that we, as evaluators, were seen as 'outsiders' and that our search for information carried a reputation of 'policing' among their network members whose work we were assessing. When we explained our problems to the programme manager of the network, he pressed us to surface the outcomes not by requesting information, but rather by asking for stories. However, this was exactly the problem. Although we sought stories, we often got stylised and idealised representations of what had happened. These accounts included claims that could not always be substantiated. However, in interrogating and analysing the narratives, we found diverse views of advocacy realities. We kept our search for information modest, establishing achievements by asking broad questions such as whether people could provide us with 'any information' that could support the stories. Still, in most cases, additional information or visible representations of the claims were not available, making it difficult to establish plausibility claims. This meant assessment was often also about assessing practical judgement in strategic intent.

This situation illustrates the negotiated nature of access to information. This experience resulted in a broader discussion about 'measurement' in which several questions were asked: What can we actually say has been achieved, considering the information we have? When do we believe a claim has been made plausible? What do we do if the information is not available? This discussion was related to another dilemma: Advocacy does not always leave evidence – an argument that was also used as a defence by the staff of the evaluated programmes.

Evidence

A key question for our team was what level of evidence was necessary. This question became significant because we learned that, although establishing proof is highly desirable, resource limitations meant that it is often out of reach. This was especially important because

international advocacy involves many actors and diverse spaces. Certainly, proving plausible attribution was sometimes possible. For example, one programme shared email exchanges with their targets that showed how the interaction led to changed phrasing in policy documents. In other cases, the levels of proof were more problematic. In advocacy, many interactions are informal and not accounted for, involving many actors and moments during long periods over which a policy agenda or plan takes shape. For many such processes in the evaluated programmes, evidence was in the form of contribution rather than attribution, and influence was often scarcely traceable, even for those who were closely involved.

We also faced the question of how to deal with evidence provided by the advocacy targets of the evaluated programmes. We were aware that these targeted stakeholders might have reasons not to be open about NGOs' levels of influence on them. In one case, the targeted organisation was not open to interviews. In other cases, the targeted policy makers interviewed were very abstract in their wording. They stressed that they were happy to work with certain programmes and organisations but that they also worked with many other stakeholders: 'Civil society has a role to play in providing content, information and knowledge on local situations which is often lacking in the New York bubble' (UN official, 2014). When policy institutions' staff changed positions, much institutional knowledge on the relationship between the advocates and the targets was lost. In such cases, the possibilities of proving influence were obviously limited. Because of such issues, discussions in our team focused on the question of whether we should aim to demonstrate causality in the change process versus the much more modest aim of establishing the plausibility of claims of contribution to change. In the end, our team negotiations resulted in the conclusion that we would seek evidence within our possibilities, which could vary across programmes operating in different contexts. We accepted that there were diverging viewpoints on the feasibility of this aim within our team, and we acknowledged that these differences could not be settled objectively in the uncertain and partly unknown situations where we did our work. This left room for those who thought they could provide evidence for causal explanations. This flexibility provided space in the team for diverse interpretations of outcomes and proof, making it possible to proceed.

Content	Number of feedback comments
Framing of text: adjustments to specific wording or arguments, stressing positive examples or perspectives, highlighting strategies	209
Outcomes specifically: including or excluding outcomes, different wording about outcomes	87
Factual inaccuracies: style, spelling, sentences	71
Total final feedback comments, 2015.	367

Table 3. Third feedback round of the advocacy evaluation, January 2015

3.4.3. Presenting outcomes

Moments of feedback around the evaluation report created space for evaluators and staff members of the evaluated programmes to negotiate which outcomes would be included in our report and how these would be included. After the first baseline report (2013), the programme representatives expressed their unhappiness about the limited space to react to drafts before the report was finalised. In response, more feedback rounds were included in the final reporting phase, providing opportunities to comment on the draft report at various stages and on the final report.

These feedback rounds were formally meant for the correction of factual inaccuracies. However, the ambiguity of the concepts of 'factual inaccuracy' and 'correction' allowed these rounds to be used for wider purposes. As Table 3 shows, factual inaccuracies were mentioned, but the comments focused mostly on the framing of the text, as well as the inclusion or exclusion of outcomes. After two rounds of feedback, the final feedback round still amounted to 367 comments taking up over 53 pages. The next paragraphs illustrate the negotiations that occurred through this process by presenting three examples: exclusion of outcomes, positive framing and adding perspective.

Exclusion of outcomes

The evaluation team found that a specific set of outcomes claimed by one of the programmes was not clearly substantiated by the realities found in the field. Through the feedback rounds, programme staff negotiated to remove this discussion from the report:

> You have to be more careful with putting criticism in the report that will be sent to the Minister. It has to be more scientific, and you have to be aware of the sources and be open to comments of the alliances. (Feedback comments, 2014)

They argued that the evaluation team was making a 'judgement' rather than an assessment. Additionally, the programme staff stressed that the data were invalid, because insufficient stakeholders, and the wrong stakeholders, had been interviewed. These staff members emphasised the need to at least 'interview all relevant people'. The team's work was criticised as being of a 'journalist style' and therefore scientifically inadequate. The programme staff stressed that the inclusion of the section in question in the report would be harmful to the sector, even going so far as to involve the coordinating agencies, expressing their concerns and putting pressure on the evaluators to exclude this finding from the report.

Although we understood the sensitivity of their position, these negotiations did not lead to the exclusion of the finding in question. For the team, the findings were accurate. The space for interpretation provided by advocacy and advocacy evaluation, combined with the pressing organisational interests involved, led to tensions and even quarrels between the evaluators and evaluated programme staff members.

Positive framing

These quarrels also happened around negotiations pressing for the findings to be framed more positively. This mostly involved programme staff pushing for stronger language about achievements and certain capacities. In the case of a women's rights advocacy programme, the negotiations demonstrated the interests at stake for the programmes, but also the complexities of advocacy evidence. This programme focused on linking local needs to global policy arenas and vice versa. Having observed local- and global-level processes, we acknowledged that the programme played a role behind the scenes as a knowledge broker and facilitator, bringing together women's representatives. Their advocacy role was not as visible in terms of policy

changes and influence. However, for the organisation, it was important to emphasise their policy-influencing role. In their feedback on the evaluation report, they requested a change to the following text:

> In this process, [the programme] plays a role behind the scenes as a broker, bringing together the important voices on specific issues and gathering information and sharing knowledge from within their own networks and partnerships. [The organisation] also plays a role as convenor […].

The organisation's programme staff wanted to change be above text as follows, framing it more strongly in terms of their advocacy role:

> Through its participation in the [policy platform] […], [advocacy] contributes actively to the Working Group's lobby on the implementation of [UN Resolution] by the Dutch Government.

In another effort to underline their identity as advocates, the programme suggested adding the outcome category 'direct lobby':

> [The programme] has done a lot of direct lobby, through engaging in one-on-one policy dialogues with policy makers at the UN, government officials in conflict-affected and donor countries, NATO officials, etc.

In their responses to the report, this programme's staff mostly positioned the meaning and role of the programme as more influential than what was found. Their comments also stressed the relation between local and international advocacy (although we did not necessarily find strong evidence for this complementarity):

> Though [country field work] doesn't focus on lobby, it is certainly complemented by our international lobby […]

Our evidence was not always comprehensive enough to validate all of the organisation's claims, as was the case for 'direct lobby'. We found that relations were built with policy makers, sometimes resulting in invitations to meetings, being mentioned as a key player, or the use and support of policy briefs by the mentioned institutions' staff members. We also found that local

and international advocacy did not clearly reinforce one another, although this was the advocacy programme's legitimacy claim for pursuing ownership in local networks, women's groups and local organisations. Evidence was limited on achievements other than taking part in meetings with targets and allies at international level and contributing to networking and local-level community initiatives.

All organisations' feedback on our evaluation report tended to add to and reframe the findings to convey a more positive or stronger role as advocates. For us as evaluators, this seemed to be an attempt to use the evaluation report as a means for advocacy – to emphasise the programme's strong advocacy capacity or demonstrate a strong influence over outcomes. While the evaluators had the final say, we had to reflect constantly on our findings, making decisions about additions or changes. This became an important aspect of reporting, because the advocacy evidence was not necessarily straightforward.

Adding perspective

The back and forth on the presentation of the findings demonstrated the sensitivities around gathering and interpreting evidence. The evaluation included case studies to get a more in-depth view of processes on the ground. To demonstrate the negotiations that happened around the presentation of these findings, we focus on a regional network that was part of a global conflict-prevention network. This network's programme management responded with concern over the choice we made to focus on the regional network:

> When it comes to documentation you may find it difficult to read about their [region] results. As I mentioned already, the region is not always good at reporting on their achievements. Not all their plans are realised in practice. Their planning is often done 'last minute'. Altogether, this is not the easiest region for you to work with […] So it is not a straightforward recommendation NOT to take [region] as a case for in-depth analysis, but we are concerned that there may not be sufficient content and results during the coming months. (program manager, 2014)

These concerns demonstrated not only the negotiation about what constituted proper data for evaluation, but also the complexity of networked processes. Proving outcomes and contribution

to changes was difficult, both because the network worked globally and because it was made up of regional offices that collaborated in the network but largely operated independently. Outcomes directly emerging from collaboration in the network were often not visible or tangible changes in policies or practices. Instead, these were contributions to societal processes such as peace education and spaces created by civil society organisations for discussing critical issues on conflict and peace.

In studying the regional network's advocacy for the case study, we observed its many challenges. These challenges demonstrated the tension between what the network pursued collectively and what it achieved regionally (see also Arensman et al., 2017a). We observed how the annual regional network meeting was used for sharing information rather than for the stated purpose of establishing a collective networking advocacy agenda and implementation plan. In discussing the agenda, members did not reflect on successes and failures or reasons for these. Instead, they considered the network to be their coming together at the annual meeting; they did not necessarily identify as belonging to the network or actively engaging with the network beyond this meeting.

When we shared our analysis, we were confronted with the organisation's discontent. They asserted that our analysis was 'weak' and 'invalid', because they interpreted the processes differently. For them, the regional processes centred on sharing knowledge and brokering contacts, although this did not necessarily result in visible changes and influence. As previously mentioned, outcomes were not necessarily the understood as changes in agendas, policies and practices, but rather as the gradually changing space for civil society to interact with policy makers, or as the space and knowledge brokered by the annual internal meetings. In their feedback, the evaluated programme staff emphasised that they had reported many more outcomes than were discussed in our draft evaluation report, negotiating for a broader, more processual understanding of outcomes.

As evaluators, we had to decide how to deal with these different interpretations and the negotiations pursued through feedback. While we considered the observed challenges relevant for the network to reflect on, we also made a practical decision to include more nuance regarding some of the challenges and limitations in the report. Revisiting the data, we tried to emphasise the layered nature of being a network. We emphasised the process and included networking-specific outcomes. We also incorporated the organisation's perspective by adding diverse quotes from partners and programme staff. However, we did not change the overall finding of tensions between what was pursued and what was achieved.

3.5. Discussion

Our findings show how effectiveness was given meaning through a collective and iterative process of identifying, measuring and presenting the outcomes. Interactions were key throughout this process. In the identification of outcomes, intangible or non-claimed achievements of multi-stakeholder processes were surfaced through dialogue and co-creation between the evaluators and the evaluated. For measuring and presenting outcomes, interactions shaped and were shaped by discussions and negotiations. Continuous negotiation between demands, access to information, the interpretation of evidence, the interests at stake and different points of view within the evaluation team resulted in valuable interactions. These interactions facilitated the development of ways to deal with diverging understandings and provided space for diverse interpretations that could do justice to the complexity of advocacy and the related challenges in its evaluation. In addition, interactions made room for learning by including multiple interpretations and confronting the evaluators and the evaluated with differences, without losing sight of the role of the evaluator as ultimately responsible for the assessment.

The interaction, however, involved more than learning through dialogue and the co-creation of understandings doing justice to multiplicity. Effectiveness was given meaning through the politics of negotiation, the demands and needs of us as evaluators (establishing effectiveness; answering the evaluation questions), and those of the evaluated program's (claiming influence to emphasize effectiveness), the provision and denial of information, and the interpretation and presentation of findings. This can be considered problematic in terms of the rising demand for measurable results pursued in the post-positivist evaluation tradition, which is concerned with prediction and control, seeking to objectify evaluation and its results.

Whereas the positivist tradition and realist evaluation assume that cause–effect relations are clear, that evidence is accessible and that outcomes are obvious, our findings demonstrate the opposite. Advocacy is often a struggle when it comes to evidence, claims, information and assessment. It is politically sensitive and shaped by relations and interactions, and the different realities of advocacy outcomes and achievements do not lend themselves to short-term measurement or predefined outcome indicators. Evaluation is a co-constructed process shaped by interactions and negotiations in which evidence and outcomes are often contested. We also found that these interactions actually contribute to evaluation quality and accuracy.

Interactions through dialogue, co-creation, negotiation, listening and engaging made it possible to adjust the preconceived notions of outcomes and measurement with which we

started the evaluation. This was highly valuable, given the need to consider advocates' own understandings of their work and their expert knowledge about their achievements. These interactions also made it possible to consider the diverse levels of evidence necessary to conduct a plausible contribution analysis, providing for a better understanding of the processes and challenges of advocacy. The challenging nature of proving advocacy effectiveness made continued negotiations over the findings unavoidable, demonstrating the highly political nature of evaluation and the multiplicities embedded in the meanings of findings. This provided space for making practical judgements and for understanding how advocacy as a practice was shaped by interactions.

3.6. Conclusion

Advocacy influences people's daily lives, as it is increasingly pursued as a strategy by development organisations, foundations, donors and corporations. Therefore, advocacy evaluation is increasingly important. Our findings demonstrate the dynamic and political nature of advocacy evaluation, where multiple realities, interpretations and interests directly or indirectly influence how effectiveness is interpreted, measured and presented. Advocacy challenges conventional evaluation methods, requiring a shift in or reconsideration of the current positivist trends (i.e. performance management, results-based management) in the development world that understand effectiveness in terms of measurable, short-term and tangible outcomes.

Although advocacy evaluation is difficult, a constructionist approach offers a way to take on the challenging issues and work with these in ways that actually strengthen the evaluation quality. Our findings show that advocacy evaluation should be conducted with a reflective, flexible approach that embraces dialogue, co-creation and negotiation. Fundamental here is an acceptance of advocacy's complex realities and of the relations and interests involved. Building on the discussions by Armytage (2011) and Ringsing and Leeuwis (2008) in this journal, this chapter demonstrates the need to open space in international development practice for reflexivity and learning about the effects of the clashing traditions of post-positivism and constructionism.

From a constructionist perspective, an interaction-based approach requires evaluators to consider their own roles, as well as the processes, relations and roles of the stakeholders, alongside the outcomes. The present analysis has provided basic insights that can be used as building blocks, rather than a full-fledged set of methodological guidelines. To take our findings

further, we are currently developing a methodology for advocacy evaluation called ‘narrative assessment’. This methodology will be tested, adjusted and further developed in practice and in interaction with practitioners.

Future research should further explore how the negotiated nature of effectiveness is shaped by politics. In our view, evaluation needs to take into account the broader funding context and existing dependencies in the development world. We make a start with this here, but we believe this matter needs broader attention. This demands a closer look into the dominant discourse around evaluation to understand the underlying roles and power relations shaping the purposes and methods of evaluation and defining evaluation outcomes.

4. Twinning 'Practices of Change' with 'Theory of Change': Room for Emergence in Advocacy Evaluation[21]

Theory of Change (ToC) is currently the approach for the evaluation and planning of international development programmes. This approach is considered especially suitable for complex interventions. We question this assumption and argue that ToC's focus on cause–effect logic and intended outcomes does not do justice to the recursive nature of complex interventions such as advocacy. Supported by our work as evaluators, and specifically our case study of an advocacy programme on child rights, we illustrate how advocacy evolves through recursive interactions, with outcomes that are emergent rather than predictable. We propose putting 'practices of change' at the centre by emphasising human interactions, using the analytical lenses of strategies-as-practice and recursiveness. This provides room for emergent outcomes and implies a different use of ToC. In this chapter, we make a clear distinction between theoretical reality models and the real world of practices.

Key words: Theory of Change; Practices of Change; practice theory; advocacy; advocacy evaluation; recursiveness; emergent outcomes; strategy-as-practice.

[21] This chapter was accepted as an article with minor revisions and published by American Journal of Evaluation as: Arensman, B., Van Waegeningh, C., and Van Wessel, M. (2017). Twinning "Practices of Change" With "Theory of Change" Room for Emergence in Advocacy Evaluation. *American Journal of Evaluation*, 1098214017727364.

4.1. Introduction

In international development, Theory of Change (ToC) is a widely used and discussed method for planning and evaluation. This method aims to address linkages between objectives, strategies, outcomes and assumptions that support an intervention's mission and vision. It began as a theory of how and why an intervention works, exploring underlying assumptions about change processes and beliefs about how an intervention contributes to these changes (Weiss, 1997, 2000). In the context of analysing how complex interventions plan for and achieve outcomes, ToC is used to acquire funding (Valters, 2014, p. 1). It is thus a management tool that is used as a formal planning document or as a broader approach to analyse how development processes work. ToC is referred to as a 'roadmap' or 'blueprint' for getting from 'here to there' (Stachiowak, 2013, p. 2; Stein and Valters, 2012, p. 3). Increasingly, ToC has become a new paradigm, touted as an ideal tool for planning and evaluating effectiveness that does justice to the complexity of development interventions (Barnes, Matka, and Sullivan, 2003; Reisman, Gienapp, and Stachowiak, 2007; Vogel, 2012).

We distinguish two applications of ToC: *ex-ante*, as an approach to strategy development and planning, and *ex-post*, as a monitoring and evaluation approach. In our evaluation of eight transnational advocacy programmes (2012–2015), we used ToC to reconstruct how change was understood, pursued and achieved (Arensman et al., 2015). By focusing on the practices of advocacy in a case study of a Pan-African child rights advocacy programme, this chapter looks at ToC primarily as an approach to monitoring and evaluation. From our evaluation, we learned that ToC, in its current form, does not do justice to the dynamic nature of advocacy practices. We asked whether the (ex-post) use of ToC could be improved to do more justice to the complex nature of advocacy, which is shown as emergent and recursive by our case study findings. In this chapter, we propose a refinement of the use of ToC, introducing practices of change in addition to theory of change.

Advocacy is an increasingly important strategy for sustainable development effectiveness. We define advocacy for international development interventions as 'a wide range of activities conducted to influence decision makers at different levels towards the overall aim of development interventions to combat the structural causes of poverty and injustice' (Arensman

et al., 2015, p. 42; following: Morariu and Brennan, 2009, p. 100).[22] Advocacy is often multilevel, with differentiated linkages across levels; multi-sited, with differentiated linkages across sites; and multi-actor, with differentiated engagements, understandings and roles in programmes, and involving multiple organisational structures, capacities and accountability relations. Although some authors claim that ToC is specifically useful for such complex processes (Beer and Reed, 2009; Jones, 2011), others have already questioned whether ToC in its current state lives up to this potential (Jabeen, 2016; Mowles, 2013).

The logic of ToC produces predefined, intended outcomes and does not sufficiently problematise the complex nature of development processes. Critics have questioned whether ToC is the right tool to deal with complex development interventions such as advocacy (Mason and Barnes, 2007; Valters, 2014). The presentation of a ToC is often a requirement for donor funding, showing the planning for implementation and for evaluating effectiveness. Some have questioned whether ToC is just the next 'trick' to perform in the name of accountability (Valters, 2014; van Es and Guijt, 2015). Others have stressed that there is a risk of ToC being imposed by the results agenda, which emphasises assessing effectiveness using predefined and intended outcomes for accountability (Eyben, Guijt, Roche and Schutt, 2015; Riddell, 2014). This raises several questions: What does this mean for the ToC approach in monitoring and evaluating advocacy interventions? Can the use of ToC be improved to mitigate the problem of unpredictability?

Based on our findings, we argue that *practices* of change should be emphasised in addition to *theory* of change. This argument is inspired by a study conducted by Teles and Schmitt (2011), who proposed evaluating advocacy by looking at the advocates rather than the outcomes. However, they did not translate this ambition into a practical approach. In response to their argument and addressing the need they identified, we bring practice-based theory, developed in management and evaluation studies, together with our case study findings. We propose using Jarzabkowski's (2005) analytical lens of *strategy-as-practice* (SAP) to give expression to practices. Using this lens provides important insight into how practices have evolved, as compared with theory. It considers strategy something actors *do* in interaction and in reaction to the organisational and contextual environments, rather than what organisations *have* as a fixed set of skills. Strategies are thus dynamic practices rather than static assets. This implies the social construction of strategies, as actors have agency over their decisions (see Long, 2003). Emphasising strategy as the product of human action and interaction makes it

[22] Advocacy comprises a variety of sub-strategies and activities, including campaigning, awareness raising, creating critical mass, lobbying and cooperating with targets you seek to influence.

possible to do justice to the complex nature of advocacy. Hence, we argue that the use of ToC as an evaluation approach can be enriched by 'practices of change'.

This chapter proceeds by highlighting the challenges of ToC as an evaluation approach. We then discuss how to look at ToC differently by introducing the discussions in management studies on SAP and recursiveness. We then introduce our case study, discussing the approach and methods used therein. Next, we present our analysis of the child rights advocacy programme to illustrate the emergent nature that advocacy outcomes have in practice. Based on these findings, we introduce a new use for ToC, taking a *'practices of change' approach to advocacy evaluation*. We conclude by summarising our findings and linking them to macro discussions in the development field and proposing suggestions for future research.

4.2. Theory of Change and Advocacy Evaluation

In international development, many planning and evaluation tools are used to gain control over programme effectiveness. ToC evolved from the linear logic model, the logframe. The logframe only lists inputs and outputs, without looking at the different elements in relation to each other. In its monitoring agenda, it focuses on intended outcomes, which is problematic for processes where not everything can be planned. Additionally, the logframe does not allow for insight into the processes leading to outcomes, or space to incorporate influences from the changing context (Prinsen and Nijhof, 2015). ToC was meant to resolve these problems.

ToC was developed from programme theory in the tradition of theory-based evaluation and is concerned with how and why an intervention works (Weiss, 1997, 2000). Programme theory studies the interrelation between the mechanisms of change, the programme and the outcomes it intends to achieve. A growing number of studies claim ToC is the best approach for dealing with complex social and political change processes, because it emphasises the links between objectives, strategies, outcomes and assumptions (Beer and Reed, 2009; Louie and Guthrie, 2007; Stein and Valters, 2012; Vogel, 2012). There is no universally agreed understanding of the exact nature of ToC, although it is often referred to as a roadmap, theory for action, or blueprint for strategic planning and learning (Reisman et al., 2007). Despite ToC being celebrated and widely used, studies have shown that tensions exist between its ambitions and its implementation, demonstrating that the approach is not yet living up to these ambitions (Eyben et al., 2015; Mowles, 2013; Valters, 2014).

For this reason, ToC has been criticised as both a planning and monitoring and evaluation tool. Mowles (2013, pp. 47–49) found that ToC is predominantly used as an outcome-based approach that evokes cause–effect thinking (i.e., *if* we do this, *then* this will be the result). Consequently, this emphasises the intended outcomes foreseen by policy-relevant theories and plans (Coryn, Noakes, Westine, and Schröter, 2011; Valters, 2014). Moreover, ToC as a monitoring and evaluation approach has triggered highly critical voices questioning whether it is the best approach for complex interventions or rather a new donor-driven results tool.

ToC follows a *cause–effect* and *if–then* logic that tends to focus on how the programme expects to achieve its intended outcomes (Vogel, 2012). Therefore, it does not pay attention to other outcomes, such as those that are not anticipated or intended.[23] Van Es and Guijt (2015) illustrated how ToC as an approach, both ex-ante and ex-post, is limited by the political pressure focusing on intended results, which does not enable a reflective and learning environment. Valters (2014) argued along the same lines, adding that, in the ToC approach, the institutional environments (i.e., bureaucracies, funding agreements and programme planning) and limited resources (i.e., time and funding) mean that accountability trumps learning. Eyben (2015, p. 21) conceived of ToC as *results artefact*, helping to achieve the politically desired outcomes that must be accounted for. This reality contradicts the ambition of ToC to be a reflective and critical approach. Thus, ToC as an evaluation approach becomes a framework for testing predefined hypotheses, as was also confirmed by Bamberger, Tarsilla and Hesse-Biber (2016).

Barnes et al. (2003) stressed that the evaluation of nonlinear processes involving multiple stakeholders and relations demands multiple theories of change. They argued that the contested interpretations of stakeholders, who all have different interests and demands, need to be addressed. Building on this, we consider advocacy a complex intervention, because it pursues multilevel change, works in multi-stakeholder settings, crosses borders and concerns diverse levels of policy arenas. Rogers' (2008, p. 39) work helps with the understanding of the outcomes of advocacy interventions as emergent rather than intended or planned. She considers outcomes to emerge through interactions during the process of implementation. Emergent outcomes, including unintended outcomes, can be mistaken for ineffective management, although these outcomes actually point to evolving, strategic and flexible approaches, as well as complex processes. If evaluators focus merely on intended outcomes, these emergent outcomes are easily overlooked.

[23] See also Jabeen (2016) for reflections on the emphasis on intended outcomes in international development discussions.

At the same time, multiple advocacy researchers have claimed that ToC is essential and strategically important for understanding the changes advocacy achieves (Jones, 2011; Klugman, 2011; Stachowiak, 2013). In our case study, the programme did not articulate a ToC specifically but was assessed on its outcomes using the ToC method. We reconstructed a ToC for that purpose. We noticed serious tensions between the cause–effect logic in the reconstructed ToC approach and the observed practices of doing advocacy (i.e. adapting to changing circumstances, communications, interactions), including recursive strategizing resulting in emergent outcomes. These tensions included the focus on intended outcomes, with other outcomes remaining concealed. There was no explanation in the ToC for how these unintended outcomes came about. Seeking to do more justice to the dynamic advocacy processes, we therefore suggest twinning theoretical models and theorised relations, as seen in the ToC, with 'practices of change', which provide space for recursiveness and emergence.

4.3. Creating Space for 'Practices of Change'

The idea of ToC adheres to the notion that planned models equal the real world. When plans are made in anticipation of change, there is an implicit understanding that change and the implementation of these plans are driven by policies. This can be overcome by emphasising and articulating practices in addition to theorising how change evolves. SAP offers an analytical lens that can make a difference in the evaluation of complex advocacy interventions.

Breaking down the existing body of knowledge around ToC, we note that 'strategy' is an important element. However, ToC is not suited to the explicit integration of strategy as an evolving practice central to change processes, but rather theorises that if we use this strategy, *then* this will be the outcome. This emphasises the intended outcomes and places the theory above changing practice. From public health intervention studies, process evaluation is another approach that pursues a shift from outcome focus to processes (Hulscher, Laurant and Grol, 2003). However, this approach also underscores the notion that planned theoretical models equal the real world while it focuses on intended processes. It does not provide space for understanding practices as strategies and interactions. Inspired by management studies, we turn this around by focusing on practices in addition to programme theorising. By looking at the processes as they happen in practice, we see strategy as something actors do while adapting to changing circumstances.

SAP is a strand of management studies that was developed from the acknowledgement that organisational processes should be understood from practice. This makes both the theorising and the practices of strategy meaningful. SAP explains how strategy is dynamic, as it is enabled and constrained by actors, human interactions, and organisational and societal practices (Vaara and Whittington, 2012). Jarzabkowski (2005, p. 7) defined strategy as a situated, socially accomplished flow of activity that has consequential outcomes for the direction and/or survival of an organisation. She showed that interactions are always pursued while keeping a future mission in mind. This implies that actors act strategically, with a mission (i.e., agenda, interest) that guides their work. At the same time, they must also deal with the daily dilemmas of reconciling a constantly changing world with the need for stability, because this is essential for an organisation to function effectively and efficiently (Jarzabkowski, 2004, p. 530). Jarzabkowski and other SAP scholars (e.g., van Wessel, van Buuren, and van Woerkum, 2011) have addressed actions and interactions as dynamic and complex, acknowledging that organisations are not necessarily a coherent whole, but rather fragmented, pluralistic and contested. SAP therefore provides a shift from performance-based approaches (in the sense of a results agenda) to actual understanding of how strategy evolves. Strategy, then, is a type of activity that is dynamic and shaped by practitioners and practices (i.e., routines, tools, methods, discourses, meetings, communication and behaviour).

Using a SAP lens enables the understanding of practices as they evolve, rather than framing them in line with a predefined theoretical model. This provides space for interaction between the theory and the practices. In other fields that bring together theory and practice, it is considered important to focus on strategies to understand how these are used to address challenges and needs, shape interactions and influence trust-building (see Coburn and Penuel, 2016). Feldman and Orlikowski (2011, p. 1249) have argued that there is a need for the development of a practice theory to enable a focus on empirical and micro-level processes constructed through relations. Relatedly, Langley (2007) has shown that bridging micro-level relations and macro environments requires process thinking. Process thinking considers 'how and why things – people, organisations, strategies, environments – change, act and evolve over time' (Langley, 2007, p. 2). SAP provides an analytical lens to gain better insight into such processes by zooming in on the practices.

As our case study will show, addressing 'how' and 'why' questions in advocacy evaluation requires an understanding of strategies as dynamic practices. Focusing on these practices provides insight on the *recursiveness* of advocacy. Recursiveness means that each decision, direction or step results from the previous ones and generates new starting conditions for the

rest of the process through interaction, learning and decision making (Crozier, 2007). This recursiveness demands the continuous reconsidering, redefining and re-strategising of advocacy strategies through feedback loops of interactions between stakeholders and organisational/environmental contexts. Through these feedback loops, new starting conditions are developed, based on learning and integrating new ways of thinking, decision making and strategising. Crozier (2007, pp. 10–13) has illustrated this point with political communications, which are interpreted through and influenced by actions and interactions. Communication and knowledge are transformed through feedback loops between the sender and the receiver, depending on how the information is received and interpreted. What happens in these feedback loops is thus largely unforeseen (Crozier, 2007). This also implies that outcomes inherent to a recursive process should be considered *emergent.* Giving space to strategies as practices enables insight into these loops and the relations between strategising, interactions and outcomes.

To date, recursiveness has not been given prominence in advocacy or in studies discussing ToC as planning, monitoring and evaluation approach. A limited number studies have mentioned recursiveness in relation to international development interventions. These studies have considered recursiveness in the context of nonlinearity, multiple pathways of change, multidimensionality and emergence. Patton (1997, p. 232) discussed recursiveness as an element of multidirectional and multilateral relationships. For advocacy, we argue that these relationships, as practices, are key to the strategies. We therefore consider recursiveness fundamental. Looking at advocacy interventions, Rogers (2008) argued that strategies need to be revisited continuously, as practices to pursue change are recursively generated through feedback loops that create new starting conditions with each output.

Whereas ToC focuses on theorising change from policies, SAP emphasises that change is initiated from practices and then should find its way back into policy. Rather than seeing processes as linear or causally linked, SAP provides space for understanding practices as recursive and their outcomes as emergent. In the next sections, we present our case study findings, which demonstrate the importance of focusing on practices to grasp the meaning of outcomes. We show how advocacy outcomes came about through recursive strategies as practices, rather than as planned change processes, as was an initial assumption in our ToC methodology. This justifies using SAP as an analytical lens for understanding the practices of change and illustrates that advocacy outcomes are emergent, not necessarily either intended or unintended.

4.4. Methods and Case Study

This study was part of a broader research project based on a transnational advocacy evaluation (2012–2015) commissioned by the Netherlands Ministry of Foreign Affairs in cooperation with the Foundation for Joint Evaluations. The evaluation was administered by the Netherlands Organisation for Scientific Research. We were involved as evaluators and researchers. The evaluation assessed eight transnational advocacy programmes that were funded by the Netherlands Ministry of Foreign Affairs and implemented by a variety of Dutch nongovernmental organisations (NGOs) in cooperation with partners worldwide. For this chapter, we drew from one of the eight advocacy programmes. This programme was the advocacy component of a broader child rights programme. It was executed by an alliance of Dutch NGOs but implemented solely by an international NGO (INGO) in Ethiopia that focused on Pan-African child rights. The Dutch alliance referred to this INGO as a 'strategic' partner.

Our research was a multi-sited ethnographic study of this Pan-African child rights advocacy programme, which we studied through time and space (Marcus, 1995). We gathered data using diverse methods (see Arensman et al., 2015 for full details). We conducted 147 semi-structured interviews and held formal and informal meetings with staff and management, programme managers and the board of the Ethiopia-based INGO. We interviewed the Dutch alliance partners and staff members of targeted organisations, including African policy makers and other African partners. We also analysed more than 200 documents, including documents internal to the organisation and external public and policy documents. We observed five strategic advocacy events in 2012, 2013 and 2014, and we gathered feedback during formal and informal meetings with staff members, individual programme managers and with the organisations' management. We conducted participant observation and outcome tracing in the Netherlands, Ethiopia, Kenya, Mozambique, Uganda, Namibia, South Africa and the United States (New York), but time and space limitations did not allow a survey of the full Pan-African scope of the advocacy programme. We were unable to visit Francophone or Maghreb countries or Central or Western African countries. Therefore, the data cannot be generalised or considered representative of the Pan-African scope of the programme.

We invested in trust-building while remaining independent outsiders to the programme. This approach gave us the opportunity to work closely with the advocacy programme's staff, enabling us to improve our understanding by gaining access to privileged information. The longitudinal nature of the evaluation made it possible to monitor how information and stakeholders developed, thus acquiring progressive insight. For confidentiality reasons, the

names of interviewees are not given in this chapter. The next section describes the case in more detail.

4.4.1. The Case Study

The African Child Policy Forum (ACPF) was founded as an organisation in 2003 in response to the need to improve the wellbeing of children in Africa. As children were not able to represent themselves in (political) processes to improve their wellbeing, ACPF stepped in as the African voice on child rights:

> We speak about the 'African Child', and not about the 'Child in Africa', as our philosophy. We are African with international values. ACPF has international value by default. We anchor our knowledge on different international standards. We have gone international Pan-African, but if we go global international we may lose our legitimate moral voice for African children, losing that specific African flavour. (Programme manager, 2012)

The organisation was based in Addis Ababa, Ethiopia, near the African Union (AU). In 2013, the organisation had 30 staff members (including support staff) and an independent Board of Trustees with international child rights expertise. Their main mission was presented as follows:

> Despite Africa's progress over the last decade and the impressive achievements to date in improving the lives and wellbeing of children, accelerated and sustained efforts are required in terms of legal reform, investment of resources and policy implementation. (African Report on Child Wellbeing, ACPF, 2013, p. xvi)

ACPF focused its advocacy work on content-driven ways to hold governments accountable in order to improve child rights and child wellbeing. In the years of the evaluation, they focused on inter-country adoption, government budgets for children, violence against children, government accountability to children and child wellbeing in African countries, amongst other issues. In theory, ACPF aimed to pursue its mission by targeting the AU, African governments and NGOs. The main strategies to bring about change were knowledge-building, speaking out against child rights violations, contributing to legal and policy reforms and effective

implementation, alliance-building of child rights organisations to forge a common voice, and collaboration with the AU and international partners. Staff members expressed a great deal of (personal) dedication to the main goal of their advocacy: improving child wellbeing. Therefore, 'Putting children on the agenda' was often emphasised in reports and interviews.

A conventional donor–recipient relation existed between ACPF and its funding partner in the Netherlands. This created power asymmetry, because the partner in the Netherlands provided funding based on a plan of action that was to comply with donor requirements.[24] The two organisations acted from different perspectives and interests (multiple realities[25]) and did not always sufficiently connect with each other.

4.5. Findings

4.5.1. Theory and 'Practices of Change'

In the evaluation methodology, ToC was at centre stage. When we arrived in Ethiopia in 2012, we were thus harnessed with concepts including 'pathways of change', 'outcomes' and 'outputs', and with tools such as contribution analysis. However, we decided to approach the programme with an open mind, enabling reflection on our preliminary findings from our document analysis.

Our preliminary study of documents and interviews with the Dutch alliance in the Netherlands had not surfaced the ToC ACPF was pursuing. The Dutch alliance partner had stressed that ACPF was mainly doing research. What advocacy or outreach was being undertaken in daily practices was not explicitly known. We therefore first tried to reconstruct the ToC and intended outcomes ourselves (a paper exercise), but interviews and documentation provided by ACPF communicated another reality as to how change was pursued and achieved in daily practices. Noting this, we questioned staff members about how they had bridged the multiple realities. Jointly, we tried again to retrieve the ToC in order to evaluate how it had worked out in practice. We understood that the notion of ToC had been introduced to ACPF by the Dutch partner. This was elaborated in a joint ToC workshop that was organised in 2012 with the goal of inspiring improved effectiveness. The staff within ACPF had accepted ToC as

[24] Further treatment of power relations is beyond the scope of this paper. Refer to Wallace (2006), Eyben (2008) and Mosse and Lewis (2005) on this issue. In the case of ACPF, the alliance partner in the Netherlands had made serious efforts to correct this imbalance, but these were not always successful (see Chapter 12 in Arensman et al, 2015).

[25] This phenomenon and the challenges of multiple realities are discussed by Hilhorst (2003).

a donor-given tool, but they had not explicitly used it in their way of working. The ToC as defined in the workshop was not really an issue of discussion with the Dutch funding agent.

Although ACPF did not have a clearly spelled-out ToC, they did think about and reflect on what they pursued and how they went about this. One of the programme managers emphasised this point:

> We do not have an overall theoretically written ToC; we have a ToC in practice, because we have internal discussions and you are always asked by the team to account for what you are doing and for the choices made, for example, on the justification of the project or the choices made. […] At our organisational level, we have a strategy document titled From Era of Rhetoric to an Era of Accountability. That gives us an overarching direction. (Programme manager, 2012)

The staff members thus implicitly emphasised the practice side of their work over the theory, explaining how they continually and recursively worked and strategised in practice. This meant that their advocacy interventions were experienced as evolving in practice rather than in a structured on a theoretical basis or institutionalised in the organisational way of working.

We also found that ACPF was, in theory, predominantly targeting its actions towards the AU, African governments and NGOs. Figure 3 demonstrates the theory through which they aspired to pursue change. Their reports, strategic documents and interviews framed this as direct and explicit advocacy towards these three targets. They often stressed how they were influencing governments:

> We constantly organise launches, which we see as advocacy events due to our identification of the country-specific issues. In country briefs, we report where a country is lagging. We have different countries to deal with, and we have local partners who take it up. Local and international bodies take it up. (Programme manager, 2012)

This interview extract illustrates how staff members spoke about their output and facilitative role as self-evidently influencing change. They had not distinguished between the immediate output of their work and the effect it had on their targets (outcome), a point that will be further discussed in the following section. Content-driven advocacy as a strategy was often discussed

as a self-explanatory cycle that naturally trickles down to results and influence: 'We are about advocacy, because whatever we produce has to have an effect, a result. We advocate for those results' (Programme manager, 2012).

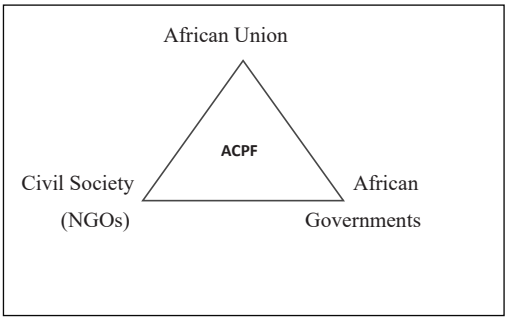

Figure 3: ACPF theory of change (Arensman et al, 2015)

In ACPF's annual monitoring reports, they spoke mainly of programme outputs, which they often presented as outcomes. We started to look beyond the output and, together with the staff, reviewed the narratives on what happened in practice to reach their targets and achieve advocacy results (outcomes). In these narratives, we found not only additional outcomes, but also strategies other than those that had been outlined in theory. Their advocacy was taking shape in a recursive way in much more strategic and emergent forms.

In addition, it appeared that governments were not a direct target for ACPF. Rather, they were targeted through the AU. Further, the country-specific activities were limited. Thus, we concluded that, in reality, ACPF followed another pathway towards change, targeting and cooperating with different stakeholders (Figure 4). Comparing Figure 4 with Figure 3 demonstrates the difference between the theory (Figure 3) and the practice of change (Figure 4) pursued by the organisation. This practice was not reflected in any theoretical model. We found that the organisation's staff members were not fully aware of the influence they wielded. Consequently, many of the outcomes they achieved were not seen, reported or followed up. In other cases, more outcomes could have been achieved if they had realised what they had set into motion with their output. In one of the countries, for example, a workshop was conducted to introduce one of the child rights reports. This workshop was highly appreciated by the participants. However, the report (recommending strengthening child rights) was never translated into the national language, and it was scarcely distributed; thus, its outreach remained

limited. The next section discusses how this advocacy work evolved and developed in an emergent way.

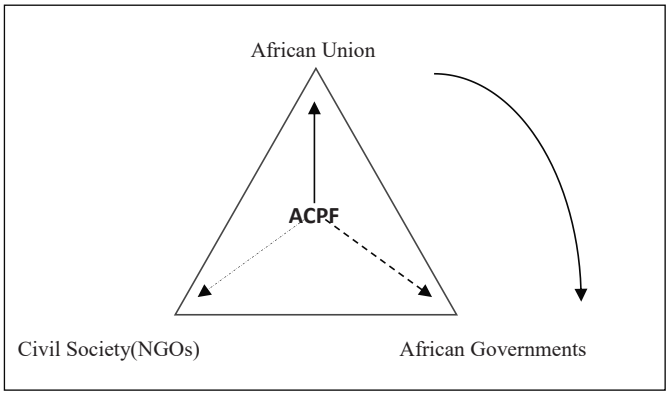

Figure 4: ACPF practice of change (Arensman et al, 2015)

4.5.2. Intended and Emergent Outcomes

In the first conversation we had with the Dutch partner organisation in 2012, it was stressed that 'their [ACPF's] research work is solid and rigorous, but the question is how they achieve effectiveness' (Evaluation manager, 2012). In other words, the work they were doing was good, but the Dutch partner did not know what the output was or what ACPF was achieving. However, upon arrival in Ethiopia, we received a document from ACPF that included their output and achievements to date. The document was also sent to the donors. It described the many activities, events and launches organised, as well as the reports written. Still, this document focused strongly on outputs, and it was missing narratives of how changes or outcomes had been achieved and of ACPF's influence.

Together with the programme managers and through country visits, we sought to co-construct the practices of change—what happened, what action was undertaken, what strategies were implemented, how and why. Through this co-construction, a different picture of ACPF's work and outcomes appeared. What ACPF had reported as achievements were often actually outputs (products of the work of ACPF's staff), but it turned out to be the case that their work had brought about many more further-reaching changes. This new understanding made it clear that many outcomes were not planned or intended but had emerged beyond ACPF's direct control or influence. Some outcomes had previously been overlooked or not understood as outcomes to which ACPF had contributed. This was confirmed particularly through interviews during the country visits.

Another overarching pertinent question arose: Why had we not learned about some of these outcomes from the funding partner in the Netherlands? We established that the result reports in the Netherlands were based on ex-ante (linear) planning (ToC) for which a monitoring protocol was designed in accordance with the donor's requirements. This raised the question of whether insufficient space and time for mutual understanding and collective reflection on outcomes between ACPF and the Dutch partner could have been a limiting factor in identifying outcomes. To phrase it differently, had this advocacy programme been conceived as a conventional development project (driven by linear theory) rather than a recursive and multidimensional advocacy intervention?

This situation, we found, illustrates the dichotomy between theory, which is formulated around intentions and assumptions, and emergent advocacy practices and their results. To understand practice, we listened to and analysed verbal and reported narratives. We learned that the challenge with advocacy is understanding outcomes as occurring in practice, rather than approaching outcomes through a ToC that starts from pre-planned results and their assessment through pre-set indicators (see Coffman and Reed, 2009; Teles and Schmitt, 2011).

4.5.3. Recursive Outcome Loops

One of the issues ACPF advocated for was 'budgeting for children', which spiralled into opportunities for exerting influence at diverse levels and layers of governance. In 2011, ACPF researched, prepared and launched a report on this issue. When we asked staff members about their achievements, they spoke about the process of publishing and launching this report. We then investigated what happened beyond this, questioning whether any changes had occurred. It was reported that the issue was brought to the attention of INGOs, which were already familiar with ACPF because they were present at the report launch in 2011. One of these INGOs kept in contact with ACPF and, in 2013, asked them to cooperate in an international consortium on child rights governance. This provided a new opportunity for ACPF to advance the issue on INGOs' agendas. The uptake of the issue of 'budgeting for children' in an international consortium generated a broader scope of interest. The consortium's collective efforts eventually resulted in targeted strategic advocacy actions aiming to influence the UN Committee on the Rights of the Child in Geneva (UNCRC).[26] This, in turn, again provided a new starting point

[26] See UN Committee on the Rights of the Child, General Comment No. 19 (2016) on public budgeting for the realisation of children's rights (Art. 4); The Child Rights Working Group on Investment in Children, UNCRC adopts a general comment on public budgeting for the realisation of children's rights.

for ACPF, because it introduced the organisation to the UN fora. An ACPF staff member explained how ACPF contributed to 'the development of a proposal to be submitted to the UNCRC' and how this evolved into a new strategy for ACPF:

> Through the Committee and focusing on child rights is the way, we can influence governments. When the Committee puts it in, we have legal pressure to influence. We submitted the proposal and made a presentation in Geneva during one of their sessions. (Programme manager, 2014)

Acting on emerging opportunities that provide for new starting conditions is a recursive loop (Crozier, 2007). The 'loop' can be described as follows: ACPF's 2011 research report led to the message that budgeting for children is required to protect children's rights. This was taken up and interpreted as something of significant value, first by the INGO and then by the international consortium and the UNCRC. This uptake became a new starting point for continuing and strengthening the advocacy message, thus expanding its influence even further:

> There is a meeting in September in Geneva, where I [a programme manager] will be talking, presenting. There is a change in relation to the Africa report and in relation to our ToC. […] We partnered for global advocacy on public spending to realise child rights in line with our 2011 report, and we wanted to take it further. This consortium provided an opportunity at various levels. (Programme manager, 2014)

This resulted in a number of new advocacy outcomes: 'The big outcome is that the Committee accepted the proposal. It is going further; they have established a working group [WG], and they asked us to provide support to this WG' (Programme manager, 2014). We also found other outcomes emerging from these recursive loops of interactions. A partnership was established around specific child rights issues. Space was created on the agenda of INGOs, and the UNCRC took up the issue of budgeting for children:[27]

---

http://www.childrightsconnect.org/wp-content/uploads/2016/08/WG_on_IC_statement_welcoming_GC_adoption_ENG.pdf (28-11-2016).

[27] This includes further cooperation with UNICEF International. See UNICEF Executive Board meeting 'Special focus session on Africa', 3 June 2014, New York, statement by Theophane Nikeyema, executive director of ACPF: https://papersmart.unmeetings.org/media2/3346480/acpf-statement-at-unicef-special-focus-session-3-jun-14.pdf (28-11-2016).

> We as [a] working group drafted the work plan, and it became a zero draft for the UN Child Rights Committee, while for us it is a final draft. The Committee looked at it, and it was accepted. And we needed a scoping document on the question—what should the scope of the document be. We are now defining the content and we knew the strategy and impact. Now our work is transferred into a general comment with key elements. And what we think should be there is there. (Programme manager, 2014)

From these narratives and reports, it surfaced that the major outcome was the influence generated at UN level for advocating for the African voice for children and the recognition that such an African voice exists. ACPF turned out to have been highly responsive to the opportunities provided:

> What has happened [is], once they [the UNCRC] had agreed, we developed a work plan on what they should be doing. We have monthly teleconferences in the consortium. We have been working very hard on it. What happened [is] there was a need to develop the working plan, and we developed it and shared it. (Programme manager, 2014)

These outcomes were emerging from interactions between the stakeholders involved. Through these interactions, information flows inspired action, cooperation and influence. This process stimulated strategising and re-strategising, which gave meaning to the advocated issues. This situation illustrates how advocacy outcomes come about through recursive practices under often unpredictable and unforeseen circumstances.

4.6. 'Practices of Change'

ToC, in its current form, is used with a focus on control and predictability. This is at odds with the need and ambition to provide space for understanding how processes develop and evolve in reality. Our case study demonstrates the need to appreciate practices in order to understand effectiveness. Our findings prove that advocacy is recursive and emergent, rather than linear and causal. In their strategic decision making, advocates interact, react and adapt. When organisations focus on what they intend to achieve (ToC), often in relation to donors'

agendas and accountability demands, they overlook what is actually achieved in practice. We suggest refining the use of existing ToCs to understand change as initiated from practices in which human interactions are central: *practices of change*. This is not another model; rather, it exposes the theoretical model to the real world of practice.

Theory-based approaches to evaluation such as ToC are valuable, because they construct (ex-ante) oversight and insight into the envisaged processes of change pursued by a programme. Critically questioning assumptions and beliefs encourages programme staff to reflect on their roles, missions, agendas and strategies. This helps to pave the way for examining the human interactions and relations. However, although it is valuable, ToC does not clarify human interactions in everyday practices. We found that the theoretical ideas about change differed from what we identified in practice, and we learned that ToC as an evaluation approach has two interrelated shortcomings. First, although it describes how and why an intervention works in theory, it overlooks practice. This results in the second shortcoming: ToC thus fails to pay attention to outcomes that were not intended. What actually happens in practice (emergent and recursive) can therefore easily be overlooked (see also Jabeen, 2016). To help us evaluate the programmes, we felt theory was one thing, but we also needed to understand practices.

The complex and unpredictable nature of advocacy demands that evaluators investigate both theory and practices. This requires a framework that provides room for both. According to Jarzabkowski (2005, p. 172), 'a framework is only valuable in relation to the conditions under which it applies.' As our case study shows, change may be strategised in theory, but it actually happens in the interactions between actors, influenced by organisational environments, policy arenas, and social and cultural contexts. This multilevel playing field is shaped by theories, practices and the interactions between the two. To understand advocacy outcomes as they are, rather than seeking to assess outcomes against a predefined theoretical framework, evaluators must examine how strategy develops in practice.

What does this mean for evaluation practice? Strategy in ToC is one theorised element (*if* we do this, *then* this happens), but it becomes central when looking at the practices—not as *if–then* logic, but rather as *how-did-it-evolve* logic. SAP foregrounds human interactions and is a way to look at advocacy processes and understand how advocacy strategy develops through daily practices. It is important to explore how such strategy is shaped over time, across multiple levels, stakeholders, borders and perspectives. Understanding strategy as something actors *do* rather than something organisations *have* gives room to practices while also making it possible to relate practice with theory (ToC). This means that it takes human interactions and processes—rather than theoretical, predefined outcomes—as a starting point for understanding

effectiveness. For evaluation, looking at the 'practices of change' through the lens of SAP and recursiveness provides a new perspective for understanding the outcomes of complex processes, such as advocacy. This includes looking at outcomes as emergent, beyond intent.

By twinning practices of change and theory of change in the way evaluators and organisations work, we believe that change, even if it happens over a long time period –which is often the case in advocacy – can be captured in terms of how it actually evolves. Twinning practices of change and theory of change thus provides evaluators and program staff with the space to follow and trace advocates' progress over time (including small steps, practices changed, strategies adapted and achievements), reflect on the practices together, and establish plausible connections between these practices and changes (even if these changes occur over a long time period). This approach thus encourages advocates to more openly record their steps as they evolve, develop and flow in their day to day practices.

In advocacy evaluation, we argue that understanding 'practices of change' demands more than questioning and theorising processes as a sequence of events that will lead to a desired outcome (Vogel, 2012) or testing and proving the theorised causal hypothesis (Mowles, 2013). This clearly has implications for the role and position of evaluators, who need to look beyond theory to observe, investigate and explore practices. Doing so will require a shift from cause–effect thinking, which emphasises outcome planning and reporting, to putting advocates at the centre. Advocates act strategically while they make practical judgements, manoeuvring through changing circumstances and acting in interaction with contexts (both organisational and environmental) and theory (plans and objectives). These practices need to be understood and reflected upon in terms of strategies, recursiveness and human interactions. Only then can an evaluator assess the broader scope of a programme's achievements.

Besides establishing common ground, evaluators should take an open, qualitative approach to collecting data. Evaluators should listen well and without prejudice, and they should critically question and investigate the strategic practices. Narratives around strategies as dynamic practices become key in establishing how change processes are shaped and developed, what human interactions are meaningful, and how this relates to diverse roles, perspectives, theories and achievements. This means that close cooperation with programme staff is necessary to create space and build the trust required for transparency (i.e., open discussions, learning and joint reflection). This demands fostering moral courage to tell an honest story, even addressing those things that failed. To optimise reflection and learning it is advisable to start evaluation of practices during the implementation.

For evaluation purposes, we consider 'practices of change' a necessary complement to 'theory of change'. Additionally, practices can be used to mirror and improve theory. This may provide plausible explanations for how outcomes (practice) relate to policy (theory) and its programme planning. The tensions between theory and practice are a well-documented field of concern in development practices (see Eyben et al., 2015; Wallace, 2006).

4.7. Conclusion

The above discussion demonstrates that ToC is potentially suitable as an approach for planning dynamic advocacy interventions, but we argue that, for evaluation purposes, ToC should be twinned with practices of change. Whereas ToC focuses on theorising change from policies, the idea of 'practices of change' emphasises that change is initiated from strategies as practices shaped by human interactions. This approach provides a practical solution to the problem posed by Teles and Schmitt (2011), who argued that current evaluation approaches are limited in terms of evaluating advocacy and suggested evaluating the advocates instead of the outcomes. The kind of twinning we suggest makes the case that theory and practice cannot be separated in evaluation.

Our approach provides the space necessary to elucidate processes that are unpredictable in nature (not predicted in any ToC), encouraging evaluators to look beyond the pursued outcomes. We argue for emphasising practices as a mirror to theories, challenging the traditions of prediction and control. Understanding the correlation between policy (theory) and outcomes (practice) clearly requires evaluators to explore and analyse connections between theories and practices. Theorising how change works provides insight into the beliefs, assumptions and ideas about change, whereas exploring how change is pursued and achieved in practice provides insight into interactions, strategies and decision making.

Therefore, we argue that the theory-based approach (ToC) should be twinned with the practice-based approach. The combination of both approaches provides useful insights into the complex processes between policy and implementation that can bridge the gap between the world of theory and the world of practices. This combined approach also provides space to take distance from and reflect on theory while doing justice to practices. New theories can be constructed on the basis of understanding practices. This should then find its way back into practice.

Our case study specifically focused on advocacy evaluation, but our findings may be relevant beyond advocacy to the broader field of working with and evaluating complex interventions that continuously evolve over time. We realise that evaluation and recording (documenting) of Practices of Change in advocacy requires further study and methodology development, acknowledging advocacy as strategic in interactions and as form of practical judgement. There is a wide body of evaluation methodologies available to draw upon also in other sectors, such as the education field. Two of the authors are also involved in developing and designing a methodology specifically for advocacy evaluation that takes a practice approach, looking at advocacy as strategic in interactions and as a form of practical judgement. We encourage more case studies to be done to accompany (advocacy) evaluations enhancing understanding of outcomes. But also to contribute to elaborating a practice-theory of evaluating complex interventions such as advocacy. Enhanced understanding and analyses of practices should also contribute to policy improvement.

5. Advocacy outcomes are not self-evident: the quest for outcome identification[28]

Advocacy outcomes are not self-evident. Identifying advocacy outcomes is extremely difficult because they are often intangible, they often arise from (personal) interactions and they are often not traceable. This challenges conventional evaluation methods. The problematic nature of advocacy evaluation has been recognised in advocacy studies, and current evaluation methods claim to do justice to these complexities. In reality, however, these methods assume outcomes are identified logically, following from actions. Based on empirical findings from a multi-sited ethnographic study of an advocacy evaluation, this chapter argues that outcomes are not self-evident, because they are constructed and political, leaving room for interpretation.

Key words: advocacy, advocacy evaluation, outcomes, outcome mapping, outcome harvesting, effectiveness

[28] This chapter was submitted as single authored article for publication in the American Journal of Evaluation and is under review.

5.1. Introduction

Evaluating advocacy is difficult. Although this has been recognised in advocacy and evaluation studies (Beer and Reed, 2009; Coffman, 2009; Devlin-Foltz et al., 2012; Gardner and Brindis, 2017; Wong, 2012), existing evaluation methods do not unpack or deal with the problematic identification of advocacy outcomes. Evaluations are often results-based, focusing heavily on outcomes and building on the assumption that outcomes self-evidently follow from actions. In this chapter, I question this assumption, drawing on findings from a quest for advocacy outcomes in an evaluation in which I was involved. The findings of the present study show that advocacy outcomes are not self-evident. Rather, in identifying advocacy outcomes, I demonstrate that there is room for interpretation, as outcomes are constructed and political. This leads to an important practical question: Should we change the way we think about outcomes in evaluating advocacy?

The relation between advocacy and effectiveness is increasingly assumed to be important by non-governmental organisations (NGOs) and funding agencies (Almog-Bar and Schmid, 2013; Hudson, 2002; Netherlands Ministry of Foreign Affairs, 2015). Advocacy has been defined as a 'wide range of activities conducted to influence decision makers at different levels' (Morariu and Brennan, 2009, 100; see also Fagen et al., 2009). Advocates pursue structural changes in social, political and organisational systems, while challenging existing power structures (Hudson, 2001; Keck and Sikkink, 1998).[29] At the same time, NGOs face growing pressure to demonstrate their effectiveness (Eyben et al., 2015; Riddell, 2014). In this context, there is a politics of results that defines effectiveness mostly in terms of assessing achieved outcomes against planned objectives (Eyben, 2015; OECD, 2010), resulting in a narrow focus on outcomes that are measurable, visible and tangible within the timeframe of funding agreements (Eyben et al., 2015; Riddell, 2014; Vähämäki et a, 2011). This involves an interaction between demands (political pressure for results) and needs (political pressure for legitimacy and credibility in demonstrating outcomes). However, measuring effectiveness is incredibly difficult with regard to advocacy (Beer and Reed, 2009; Devlin-Foltz, 2012). Practitioners and evaluators working with advocacy continuously struggle to come to terms with the demand for demonstrable effectiveness and with the complexity of evaluating advocacy. This makes research on advocacy evaluation ever more important as a field of study.

[29] This broad definition suits the wide range of approaches implemented under the umbrella of advocacy, including awareness raising, campaigning, creating critical mass, blaming and shaming and lobbying.

The complexity of advocacy is widely recognised. Advocacy studies that focus on effectiveness and evaluation point out the non-linear and political character of advocacy (Jones, 2011; Keck and Sikkink, 1998; Teles and Schmitt, 2011; Wong, 2012). Advocacy involves processes that are unpredictable and conducive to change in the highly complex sphere of public policymaking, with many actors and factors influencing the course of events (Grantcraft, 2012; Guthrie et al., 2005; Reisman et al., 2007; Roche and Kelly, 2012a, 2012b). A recent study by Arensman et al. (2017b) demonstrates the recursive character of advocacy and the emergent nature of advocacy outcomes. Many forces are at play that make or break an advocacy effort, thus influencing advocacy effectiveness. Determining the effects of strategies ahead of time in the planning process is thus likely to be impossible (Coffman and Reed, 2009; Devlin-Foltz et al., 2012; Teles and Schmitt, 2011). This means that the determination of outcomes is not 'set in stone', challenging evaluators to search for other ways of demonstrating the effectiveness of advocacy (Jones, 2011; Teles and Schmitt, 2011).

The recognition of the complexity of advocacy has resulted in important efforts to develop more suitable evaluation methods for complex interventions. Some scholars have pointed out the need to evaluate advocacy by focusing on the advocates' skills and capacities to act strategically (Arensman et al., 2017b; Teles and Schmitt, 2011). Others have stressed the need to assess contributions to changes, rather than attribution (Coffman, 2011; Gardner and Brindis, 2017; Mayne, 2008, 2012). Increasingly, scholars, practitioners and evaluators promote working with a theory of change to increase the flexibility of planning, strategizing and assessing advocacy programmes and other complex interventions (Beer and Reed, 2009; Gardner and Brindis, 2017; Klugman, 2011). This approach often includes looking at advocacy progress in terms of interim outcomes instead of major policy outcomes, which are not always achieved (Coffman, 2011).

Although the literature has widely recognised the challenges faced in evaluating advocacy and sought better ways to evaluate advocacy, the identification of advocacy outcomes remains understudied. Outcomes are generally considered central to the evaluation purpose of demonstrating effectiveness, and they are usually considered to be self-evident. The complexities and challenges of non-linearity and unpredictability have not necessarily led to the development of methods appropriate for advocacy evaluation. This chapter discusses two methods that are currently widely used in advocacy evaluation: outcome mapping (OM) and outcome harvesting (OH) (Jones and Hearn, 2009; Wilson-Grau and Britt, 2012). Despite their popularity, these methods do not provide enough room to fully comprehend the problematic nature of advocacy outcomes, because both methods take outcomes for granted.

To explain the limitations of these methods, this chapter introduces findings from the analysis of a quest for outcome identification in an advocacy evaluation. These findings demonstrate that advocacy outcomes are not self-evident, but rather interpreted, constructed and political in nature. The chapter explores the question of whether researchers, evaluators and advocates should change the way we look at outcomes. The findings in this chapter are based on an in-depth study of an advocacy evaluation in which I was involved as both a researcher and evaluator from 2012 to 2015. The evaluation was set up to assess the outcomes of multiple advocacy programmes, and the evaluation team invested a great deal of time in identifying outcomes.

The chapter begins by discussing the current state of the art of advocacy evaluation. Next, I explain the scope of the case study, the data collection and the analysis. I then discuss the findings, providing illustrations of the quest for advocacy outcomes and the problematic nature of identifying these outcomes. Finally, I conclude by drawing out the broader implications of these findings for theory and practice regarding advocacy evaluation.

5.2. Evaluating advocacy outcomes

Coming to terms with challenges associated with demonstrating effectiveness is an ongoing struggle in advocacy evaluation practice and in research on advocacy evaluation. A crucial issue is the problematic nature of advocacy outcomes and their identification. In the complex world of public policymaking, outcomes are mostly invisible, unpredictable and dependent on many actors and factors (Whelan, 2008). The relations between the advocacy efforts, progress towards outcomes, and actual changes are not easily established; these links are elusive and often indirect (Teles and Schmitt, 2011). Because they come about as a result of often intangible processes that are not observable or traceable, advocacy outcomes are both unique and problematic.

The advocacy evaluation literature discusses the problematic nature of outcomes in a number of ways. Some scholars have stressed the impossibility of establishing attribution and emphasised the need to evaluate advocacy outcomes by looking instead at contribution (Devlin-Foltz et al., 2012; Fagen et al., 2009; Jones, 2011). The importance of understanding different layers of outcomes, including short-term or interim outcomes, has also been stressed. Looking at interim outcomes rather than end goals is important because the end goals are usually abstract (e.g. the prevention of violent conflict) and are often not achieved within a specified timeframe (Coffman, 2009, 2011; Harvard Family Research Project, 2008). Specifically, interim outcomes

are the smaller steps taken towards a bigger change. Types of interim outcomes include changing attitudes or beliefs, raising awareness, creating visibility, building relations and creating space for civil society. These steps might not result directly in explicit changes in policies, laws, practices or institutional structures, but they are nonetheless meaningful. Focusing on interim outcomes, some have claimed, provides the space necessary to demonstrate changes, following and assessing progress towards larger pursued objectives (Devlin-Foltz et al., 2012; Harvard Family Research Project, 2008; Starling, 2010).

Although these problems in advocacy and evaluation studies are acknowledged in the field, the advocacy evaluation approaches that are currently widely implemented are not necessarily helpful for gaining a better understanding of advocacy outcomes. Approaches like OM and OH are 'hot' methods in the NGO community, implemented with the intent to do justice to complex processes such as advocacy (Jones and Hearn, 2009; Wilson-Grau and Britt, 2012). The methods, however, do not problematise outcome identification as such. They build on the assumption that outcomes can be predicted and measured, or at least that they are self-evident. OM, in the tradition of logic models and theory of change frameworks, clarifies the presumed logical intended relationships between objectives, predetermined outcomes, activities and achievements (Smutylo, 2005). OH is an evaluation approach that does not measure progress, but rather works backwards, collecting evidence on the outcomes achieved (Wilson-Grau and Britt, 2012). Both approaches are considered relevant for evaluating advocacy.

Outcome mapping

OM was developed in 2001 (Earl et al., 2001). Jones and Hearn (2009) have characterised the method as a people-centred approach to planning, monitoring and evaluating change. This approach claims to acknowledge flexibility, complexity and non-linearity in its recognition of a 'complex web of interactions between different actors, forces and trends' around outcomes (Jones and Hearn, 2009: 2; Earl et al., 2001; Smutylo, 2005). OM considers outcomes to be changes in behaviour, relationships, networks, or the actions and activities of individuals, groups and communities. To map outcomes, a high degree of cooperation with those evaluated is necessary, and a shared vision on what is being evaluated and mapped should be established (Tsui and Lucas, 2013: 7). Three stages are specified as important: 1) the identification of intention (how, why, who, what), outlining the programmes' desired change; 2) monitoring performance to identify progress towards outcomes and goals; and 3) evaluation planning to identify priorities.

The theory of OM considers outcomes as progress steps rather than necessarily being big changes. Earl et al. (2001: 14) have stressed that OM deals with the challenge of attribution by focusing on the value and intention of results achieved: 'It does this by focusing on the changes that are clearly within a programme's sphere of influence. While, at first glance, this appears to suggest concentrating on easier, less important, short-term achievements, in fact it does the opposite. It focuses attention on incremental, often subtle changes, without which the large-scale, more prominent achievements in human well-being cannot be attained or sustained.' OM thus takes interim outcomes into account, but, in doing so, it also considers outcomes to be self-evident. Tsui and Lucas (2013) have highlighted that OM works well when outcomes are clear from the start. This follows from the method's assumptions that outcomes can be identified and are observable.

Outcome harvesting

OH looks like OM in that it focuses on outcomes specifically. However, rather than mapping outcomes progressively, OH harvests the outcomes achieved. OH is primarily an evaluation tool for identifying, formulating, defining and making sense of outcomes. OH is designed to help 'harvest-users' focus on what was achieved and how. The process is focused on designing the plan, gathering evidence, engaging with involved stakeholders, substantiating the evidence, and analysing and interpreting the findings (Wilson-Grau and Britt, 2012). OH claims to do justice to complexity because the method considers a highly participatory process necessary, where 'harvesters' facilitate and support the participation of those being evaluated to ensure a credible harvest. The information is collected from the individuals or organisations implementing the evaluated programmes.

In OH, outcome descriptions are specific (formulated in detail), measurable (independent and objectively verifiable quantitative and qualitative information), achieved (plausible relation and logical link between the outcome and the agent's actions), relevant (the outcome presents a significant step towards the main goal) and timely (the outcome occurred within the evaluation period) (Wilson-Grau and Britt, 2012: 8-9). The evaluator is meant to 'harvest' the verifiable connection between the initiative and the outcomes, creating outcome descriptions that answer the harvesters' questions and revising these descriptions to make them more specific and comprehensive. The descriptions of the harvested outcomes are then validated 'by comparing [them] with information collected from other knowledgeable and authoritative, but

independent, sources'[30]. Finally, these outcomes are analysed and interpreted in relation to their significance for achieving a certain mission, goal or strategy.

OH focuses on progress and behaviour change. The method concentrates on collecting evidence of achievements rather than measuring progress towards predetermined outcomes. The evaluator determines the contribution to an outcome by collecting data from reports, documents, interviews and other resources (Lemon and Pinet, 2017).

5.3. Limitations of the existing methods

Although OM and OH are widely used to evaluate and plan advocacy (Lemon and Pinet, 2017; Roche et al., 2012; Tsui et al., 2014; Tsui and Lucas, 2013; van Ongevalle, 2012), they do not contribute to solving the problem of how to identify unobservable, intangible or nonlinear advocacy outcomes. Both methods have several critical limitations when it comes to really doing justice to advocacy as a complex practice. In the context of increasing pressure on NGOs to demonstrate effectiveness by showing measurable and visible results, the trend is to focus on quantity rather than quality (Best, 2017; Eyben et al., 2015; Natsios, 2010; Riddell, 2014). OM and OH explicitly focus on visible outcomes and do not question the problematic identification of outcomes (Smutylo, 2005; Wilson-Grau and Britt, 2012). For this reason, both methods may overemphasise the assessment tangible outcomes in light of what was planned, rather than understanding the process of how these things came about.

The two approaches assume that evaluation has objectivity, but this premise can be questioned. The methods build their outcome descriptions and findings based on what is reported by those implementing the evaluated programmes, validating these claims through other 'knowledgeable and authoritative' sources of 'independent' information (Betterevaluation, 2017; Smutylo, 2005; Wilson-Grau and Britt, 2012). This approach implies the assumption of the existence of independent knowledge that will allow for the objective evaluation of outcomes. However, previous research has demonstrated that evaluation itself and the role of outcomes are both political (see Arensman et al., 2017a+b; Taylor and Balloch, 2005) and that evaluators of advocacy have to build their assessment on partly subjective and incomplete information (Teles and Schmitt, 2011). The methods that are currently widely used

[30] From Betterevaluation (2017), source: http://www.betterevaluation.org/en/plan/approach/outcome_harvesting quote can be found under point 5 'How is outcome harvesting done?'.

do not take these issues into account or provide any guidance to practitioners working encountering them.

Both OM and OH fail to discuss the roles played by politics, demands and needs (claiming, visibility), power inequalities and dependencies (funding, access to information, resources), and how all of this affects the evaluation results. Despite being 'hot' in international development, these two methods have critical limitations. As my findings will demonstrate, especially in advocacy, outcomes need to be surfaced in interaction with the stakeholders involved, and this presents critical challenges to the existing methods discussed here and their assumptions about outcomes. The two methods should be adjusted based on a better understanding of advocacy and its outcomes.

5.4. Case studies, data collection and analysis

5.4.1. Advocacy for development evaluation.

This study is part of a broader research programme conducted from 2012 to 2015 focusing on an evaluation of international advocacy programmes. The programmes were financed by the second Dutch co-financing system (2010–2015), known as MFS II.[31] The evaluation was commissioned by the Netherlands Ministry of Foreign Affairs (NL MFA), in cooperation with the Foundation for Joint Evaluations, and administered by the Netherlands Organisation for Scientific Research. I was involved as evaluator in a team of 10 evaluators and as a researcher conducting my PhD research.

The evaluation assessed eight advocacy programmes funded by the NL MFA and implemented by a variety of alliances of Dutch NGOs in cooperation with partners worldwide. For this chapter, I draw on three advocacy programmes that I was closely involved in evaluating. For confidentiality reasons, I anonymised the case studies and the interviewees. The three programmes were an NGO working on child rights at the Pan-African level, an international NGO working on women's rights at local and global levels, and a transnational advocacy network on conflict prevention and peacebuilding. In all cases, the advocacy was conducted in cooperation with global, national and regional partners.

[31] In Dutch: *Medefinancieringssysteem* (MFS II), MFS II ILA Evaluation. ILA stands for International Lobbying and Advocacy.

The main purposes of the evaluation were to account for the results of the funded advocacy programmes, to contribute to the improvement of future development interventions and to develop advocacy evaluation methodology. The evaluation team was tasked with addressing evaluation questions that were predefined by the donor. These pertained to the outcomes achieved, contribution, relevance and efficiency of the programmes, as well as explanatory actors and factors. The team was instructed to answer these questions looking specifically at three predefined priority result areas: agenda setting, policy influencing and changing practice. These result areas supposedly mirrored an expected sequence of outcome areas (not necessarily in a linear fashion) of the advocacy programmes.

5.4.2. Data collection and analysis

My research was a multi-sited ethnographic study of the three advocacy programmes and of the evaluation process itself. I studied the advocacy programmes through time (2012–2015) and space (at local, regional, national and global levels). I participated in all 15 internal evaluation team and management meetings, in which decisions were made regarding the evaluation direction, strategies and methods. As part of the evaluation team, I collected data on the three programmes discussed in this article via semi-structured interviews, document review, participant observation of the advocacy programmes (in the Netherlands, Ethiopia, the Philippines, New York and Ghana), 29 meetings relevant to the advocacy programmes, and three meetings relevant to the evaluation process (including all stakeholders). My co-evaluators and I conducted interviews in the three programmes in Kenya, Mozambique, Uganda, Namibia, South Africa, the Philippines, Ghana and New York (see Annex 2 for an overview of the interviews). Interviewees were NGO representatives from the evaluated programmes and other stakeholders such as their international and domestic partners, targets and other relevant experts in the field. I reflected on and analysed the processes based on all of the data collected (interviews, observational field notes, written documents, emails and team discussions).

In my position as an evaluator and academic researcher, I was an insider (evaluator), an observer (evaluator and researcher) and a researcher. While I was working as evaluator in the evaluation team in my capacity as an evaluator/researcher, I observed the processes and interactions and was able to reflect on them. Wearing different hats in this way gave me a unique opportunity as a researcher: I was able to work closely with the advocacy programmes' staff and to research the evaluation as a process. This enabled access to privileged information that

allowed me to acquire useful insight, and the analysis is informed by my dual position as both evaluator and researcher.

5.5. The quest for advocacy outcomes

The following section demonstrates three problematic dimensions of the quest for identifying advocacy outcomes. First, outcomes become outcomes when they are interpreted as such by stakeholders making (strategic) decisions. Second, advocacy outcomes are constructed in the way they are given meaning in the multi-layered realities that influence how they are invoked and presented. Finally, the process of identifying outcomes is political, as outcomes are given meaning in relation to diverse interests and (strategic) decisions made regarding needs (legitimacy, credibility) vs. demands (accountability). This section underscores the complexity of advocacy evaluation and provides a basis for critical questions on whether we should let go of outcomes as the standard for effectiveness. This is directly linked to the question of measuring advocacy effectiveness (what vs. how).

5.5.1. Room for interpretation

The findings of the advocacy evaluation revealed that outcomes have different meanings depending on the interpretation of the advocates and on their aims. Something considered an outcome for one person may not be considered an outcome for someone else. Outcomes become outcomes through the interpretation of the stakeholders. This is dependent on the interactions, interests and contexts involved. Especially in advocacy, there is room for interpretation around outcomes, because processes, outcomes (in terms of change achieved), and the relation between advocacy and the changes achieved are not obvious. To make this explicit, I discuss the cases of claiming and non-claiming and of underreporting outcomes, revealing the space for interpretation as to what an outcome was and how its meaning was used.

Claiming and non-claiming

The evaluated child rights advocacy programme was part of an organisation with a diverse staff that was very committed to changing the lives of children throughout the African continent. Early on in the evaluation, we noted that the organisation and staff had a very good reputation with their targets and partners; they were considered knowledgeable on the topic of

African child rights. However, the outcomes identified in reports or claimed by staff members in interviews or meetings were limited. To sketch the context, the organisation was situated in Ethiopia, close to the Pan-African political arena of the African Union (AU), and they worked closely together with AU, African governments, the United Nations (UN), and other civil society organisations and networks. Our evaluation focused on advocacy and influencing, but this organisation's staff members did not see their work in this way. Because they were based in Ethiopia, they had to consider the country's restrictions on NGOs' roles (i.e. advocacy was not permitted) and funding (limitations to foreign funding).

The organisation had built credibility from their years of doing research, planning events, working closely with diverse public and political stakeholders on the African continent, and advocating for the issues identified as gaps, or problems in national and regional policies on child rights. The role of the organisation was pointing out the issues they surfaced through their research to the stakeholders involved to raise awareness and devise solutions to existing problems. The organisation could influence the awareness of their partners and targets by doing research to point out problems and by disseminating this research, translating it into diverse African languages and organising workshops and conferences. However, the organisation did not have control over what other stakeholders would do with the reports they produced, the gaps they identified or the solutions they proposed.

The organisation did not interpret processes or outreach over which they had no control as outcomes. These outcomes were therefore outside their view of direct control, and they had not thought of them as being relevant strategically. These outcomes were thus not identified or claimed. One programme manager described their work as 'a chain of action' in which their research was the basis for further influencing:

> 'Some is reactive and most of it is pro-active. Like […our] campaign. This started in January when we knew the department of social affairs was going to focus on [the issue of our campaign]. We are looking at this from a legal and rights perspective, because that is what the AU is looking at. We share research papers with the [inter-governmental department]. From the discussion based on the research we had done, we were asked to help draft a declaration […] and it was discussed and adopted (11 April 2014). We wrote the declaration, they asked us, but we cannot say that outside.' (Programme manager 2014, Addis Ababa).

In this case the advocacy target took up the declaration written by the organisation, and put it on their website as theirs. This in itself was an outcome that could be attributed to the organisation's work and good credible reputation, but it also illustrated their way of working strategically with targets. One of the program manager's explained that much of their work was pro-active trying to find gaps to jump into and align to target campaigns or processes. The programme manager said:

> 'We want to inform the [target], so we look at [specific] policies tailored towards supporting the [..] campaign to add value to their [the target's] process. It is not an abstract research. Research directly geared towards an outcome that will be directly influencing [the target's] activities or CSO activities.' (Programme manager 2014, Addis Ababa).

The advocacy target's campaign was built on information, documentation and advocacy work prepared and supported by the organisation in Ethiopia. This was underlined in interviews with the target and other stakeholders, however the outcomes were not reported or claimed. Most AU interviewees commended the organisation for their cooperation based on expertise, knowledge driven content and for its position in the background. Interestingly, the organization was explicitly lauded for not claiming ownership over documents they wrote, researched or co-authored. They organized and facilitated meetings without claiming the honours or putting up multiple banners and logos on their work or documentation. Specifically the AU officials mentioned this to be important for their good working relations. The organisation raised issues to the AU agenda without attaching their name to them. The fame and acclaim of the organisation could therefore be attributed in part to the non-claiming of what they delivered to others (outcomes); ownership was left where it belonged.

For this organisation, in most cases, non-claiming was a deliberate choice to cooperate closely with their targets. Because the organisation chose not to claim their work directly—and thus also engaged in the non-claiming of outcomes—targets were more likely to ask for input, cooperate on specific issues and extend invitations to meetings. Processes were strategically steered by cooperating closely with the main targets. One of these roles, the organisation explained, was translating every report into multiple languages spoken within the AU. This was that part of their advocacy on the Pan-African level, done to ensure that their message was conveyed to all AU member states. However, the non-claiming of outcomes also resulted from overlooking some outcomes.

To surface these kinds of outcomes, as evaluators, we had to rely on interviews to reconstruct what happened, how one thing led to another and who was involved in what capacities. This meant that there was room for interpretation—not only because the organisation interpreted their own role as supporting and facilitating rather than as advocating or influencing, but also because the information from the interviews was built on how the persons involved interpreted the events and their roles and actions. The good reputation and good working relations of the organisation allowed us, as evaluators of this organisation, to gain good access to key people in diverse (policy) arenas and civil society actors in Ethiopia and in other African countries (Mozambique, Uganda, Kenia, Namibia and South Africa). In interviews with policymakers, the organisation was lauded as 'knowledgeable', 'real experts', people who 'know what they are talking about', 'very helpful' and 'the only African voice on these issues'. These views, of course, are also interpretations.

Underreporting

In another case in the evaluation, we uncovered the underreporting of outcomes in a peacebuilding and conflict prevention network. The network committed a great deal of their global secretariat's time to collecting and understanding outcomes achieved by their regional members. Their annual reports included their members' reports on what was achieved. Often, the network's reported outcomes stated that internal network agendas were set, meetings were held, or the network facilitated learning and knowledge-sharing by bringing the members together regionally and globally. Although these outcomes were significant for what the network was generating and developing, it was less clear what they had achieved as a network or how the outcomes came about.

Setting a network agenda during an annual meeting, including action plans for the following year, was often described as an important outcome. In one of the regions, I spent four weeks observing the regional processes and three days observing the annual meeting. This meeting brought together six network members from countries in the region, three members of the regional secretariat and a representative from the global secretariat. The meeting was supported and facilitated by the regional secretariat of the network. Prior to and during the annual meeting, the members were facilitated to meet with inter-governmental organisations to discuss regional issues and with other civil society stakeholders and organisations involved in a regional conflict. During these meetings, the members were surprised to be able to meet with so many important stakeholders involved in the conflict and peace processes.

Many of the discussions at the annual meeting were about the regional network agenda. This agenda included the reported outcomes of the previous year. In these reported outcomes, the annual meeting was considered an outcome in terms of 'setting the regional agenda', making 'action plans for the following year', and 'knowledge-sharing and learning from each other'. For many of the network members, this meeting was one of the few moments that they felt was a 'network activity'. They did not interpret the work they did within their own organisations as being part of the global or regional network.

The discussions held during the annual meeting underscored that, although the reported outcomes were indeed network-strengthening outcomes, the previous year's action plans had not been followed up on by the individual members or with action as part of the global network. Additionally, the members did not report on outcomes or actions undertaken related to being part of the global network or to the advocacy agenda. However, updates on the local and regional contexts provided during the meeting made it clear that there were outcomes of these types to be reported. The interviews I conducted with network members also highlighted local-level outcomes such as the development of local networks inspired by being part of the regional network and multi-stakeholder meetings inspired by shared experiences. The network members were not aware that these were outcomes or they were uncertain regarding the relation of these outcomes to their role in the global network.

Although a regional network's agenda formulation could signal a change, it was uncertain for us as evaluators whether this could always be considered an outcome. This was a case of over-reporting internal network outcomes while underreporting outcomes that were the result of belonging to a global network (i.e. not being aware of these outcomes or not sharing the same understanding of their meaning).

The issues of claiming and non-claiming (i.e. underreporting) advocacy outcomes demonstrate that there is room for interpretation in the identification of outcomes. Outcomes became outcomes when invoked or claimed as outcomes by the stakeholders involved. In this process, diverse steps were taken through which meaning was attributed to outcomes based on the decisions of political and organisational considerations about the values and needs for partnerships, access, trust-building and legitimacy. The first organisation discussed in this section, by not claiming outcomes or its ownership of them, in combination with its content-driven advocacy, gained a widespread reputation for being a Pan-African expert organisation. This provided them with access to be able to work closely with their targets. Non-claiming provided the necessary space to build trust, be seen as a legitimate voice and establish cooperative working relations through which outcomes were achieved. For the second

organisation discussed here, however, underreporting outcomes demonstrated that they were not aware of and did not share the same understanding of the meaning of outcomes that resulting from belonging to a global network. It was thus also possible for non-claiming and underreporting to result from failing to make claims or identify outcomes, and this also meant that the organisation did not reflect upon the outcomes. This lack of awareness of their outcomes led to concerns among the organisations' donors and others about the organisations' effectiveness.

5.5.2. Outcomes are constructed

Advocacy outcomes are constructed through interactions among diverse actors working in varied contexts. This means that outcomes become outcomes through interactions among actors across multiple layers of social, organisational and cultural realities. These actors create outcomes through recursive and strategic actions and decisions. These constructed outcomes are often built over multiple years; for example, this is true of outcomes regarding the nurturing and strengthening of relations of trust and reputation. Meaning is given to advocacy outcomes in the construction of social, organisational and cultural realities resulting in multiplicities of meanings. This section shows the constructed nature of advocacy outcomes, which makes it difficult to answer questions about 'which outcomes', 'whose outcomes' and 'what outcomes'.

Annually, New York hosts multiple international weeks on women's rights. During one of these weeks, the 2014 Commission on the Status of Women (CSW) week, which is dedicated to the promotion of gender equality and empowerment, I followed and observed a number of advocates. These advocates and their partners were in New York advocating for a global fund for the implementation of United Nations Security Council Resolution (UNSCR) 1325, which pursues women's rights and equality. The advocates I followed worked with a Dutch international NGO in cooperation with other advocates through networks and organisations in diverse policy arenas (e.g. the United Nations, the European Union and the North Atlantic Treaty Organization) and contexts (priority countries including Afghanistan, Burundi and Colombia). The advocacy was directed at pursuing women's rights and women's leadership in local, regional and global contexts including the United Nations (UN). During the years of my observation and research of the advocacy programme (2012–2015), the advocates followed the orbit of global policy meetings, for example, making multiple trips to New York each year. The business of these trips comprised meetings with partners, networks, civil society stakeholders

and policy makers within the UN and with government representatives to the UN based in New York.

During the CSW, the advocacy programme facilitated and organised country-themed events, and they used this time to meet to discuss strategy. For these meetings, they invited their local partners. The local representatives to the 2014 CSW were from South Sudan, Afghanistan and Somalia. All of these representatives were women. Most of the women had been to New York multiple times. One of the representatives from Afghanistan said that she felt responsible for the future of her country and for Afghan women. She emphasised her reasons for coming to New York as telling 'the true story, which is not pretty, but I draw it out to them [the UN] the way it is' (advocate 2014, New York). With direct and indirect support from the Dutch international NGO, much of the advocacy by the local partners was driven by personal commitments to improve the situation for women in their countries. These local voices were incredibly important for legitimacy reasons and for constructing the advocacy messages. During a breakfast meeting, a Dutch advocate proposed that key recommendations from the local partners be put in a policy brief that was being distributed during the CSW week in meetings with key targets.

During the CSW week, many meetings were held, including a meeting with a high-ranking, established UN official. The advocacy organisation identified this meeting as an outcome. The advocates involved in the meeting stressed the importance of this meeting: 'If she [the UN official] underscores appreciation for our idea [global fund for UNSCR 1325 implementation], this is so high up the chain that even UN staff will see the importance of the issue we advocate for and see this as an opportunity' (advocate 2014, The Hague). While the commitment of this UN official was necessary for the process to continue, the meeting and its outcomes involved many actors, interactions and political sensitivities. For the global fund for UNSCR 1325 to succeed, the advocacy organisation needed governments on board, and the organisation's local partners at all levels were needed to contribute to getting the governments on board. The organisation needed commitments in terms of political agendas and funding. The meeting with this UN official was a constructed outcome, as it represented a spiralled opportunity emerging from years of interactions initiated by many different partners in many diverse arenas. Years of relations-building went into this outcome, as did lots of legwork, pushing and pulling, and creating visibility. Processes that include bureaucracies, such as those found in the UN, demand patience, much going back-and-forth and many follow-up meetings to keep the issue on the agenda.

Prior to the CSW week, a Dutch advocate explained to me that there was a lot of 'legwork' on the part of their partners in New York—meaning many points of contact (i.e. emails, meetings and calls) with those involved in the decision making processes, partners in the civil society arena and partners in the policy arena. This advocate stressed that over 100 emails went back and forth to establish meetings, share reports and create visibility in order to profile themselves as 'thought leaders' (policy officer 2014, The Hague). At the heart of the advocacy was 'building and nurturing the relations with civil society networks and policy makers' (advocate 2014, New York). Most of these relations became personal relations over the years.

This also included close relationships with targets, who were often also involved in the networks. A partner advocate working in the UN arena and the women's rights field for over 20 years stressed that close relations with targets were critical for their advocacy:

> Many donors want their partners to become stars because it draws also attention to their work and their role. But there are limited resources and donors have their priorities and it is a continuing challenge to reconcile donor priorities with our own priorities. Luckily, we have been recognised, and we don't need to shift all the time to the sexy issues of donors. But the funding is short-term [and] often annual, so this is a problem. (advocate 2013, New York)

A Dutch advocate from the evaluated program added to this that, '[a]s an NGO alone, you can't do it, as it is about global issues; you need governments and the UN to back you up. CSW provides strategic opportunities.' She explained that doing advocacy in this context means 'continuously pushing and pushing' (advocate 2014, New York). Another one of the advocacy organisation's partners I interviewed during the CSW week in New York stressed the following:

> The traditional civil society stance to the government is adversarial, seeing government as an enemy. For us, we can't be stuck in this traditional adversarial stance. Government bodies and elected bodies—we need them at our side. We are critical, but we cannot speak openly to criticise them. We need to build and nurture our relations, based on a good track record and being respectful while giving them space and being in constant contact.(advocate, 2014, New York)

For the organisation's advocacy, the local embedding of these voices, the organisation's presence and visibility as a thought leader, and the relations built with targets on political and social levels were extremely important in the construction of outcomes.

The above demonstrates that advocacy outcomes were constructed. First, the outcomes were layered, spiralling out from years of relations-building and connections on personal and professional levels forged between the actors involved across social, cultural and political levels, including work with both civil society and advocacy targets (policy makers at national and international levels). Second, following from this process, the outcomes were constructed on multidimensional levels involving transnational processes and crossing national borders and international arenas; they involved multiple stakeholders working at diverse levels of organisations, governments and institutions focusing on national, international and regional processes. Third, this meant that the outcomes were constructed in the sense of involving cross-political levels, with the advocacy targets themselves being actively involved in the networks. These characteristics made the changes achieved more elusive and more difficult to articulate.

5.5.3. The politics of outcomes

In addition to outcomes being interpreted and constructed, identifying outcomes is also political. There is a need for organisations to account for outcomes to prove the effectiveness of changes to which they contributed and also to establish a certain legitimacy and credibility with other stakeholders in the field and with the donors. The following section demonstrates how the politics of outcomes manifests itself in the claiming of visibility and the over-reporting of outcomes.

Visibility

During the advocacy evaluation, the question of 'whose outcomes are we talking about?' came up multiple times. Most advocacy programmes worked with local partners in networks or partnerships. In one case, it became clear that the question of 'whose outcomes' was related to making outcomes visible to claim them as your own. In 2014, as evaluators, we met with the staff of the evaluated programme for a discussion of their advocacy challenges relating to connecting global-level processes with local needs and vice versa. During this meeting, advocates and programme managers stressed that visibility was important for reasons of organisational accountability and legitimacy. The programme needed to make certain results

visible to account for funding spent (organisational) and to build credibility as a 'thought leader' (advocacy).

In their advocacy, the advocates pursued ownership. Ownership by (local) women's groups was emphasised, stressing that 'what we do is really working with women's networks' (programme manager, 17 February 2014, The Hague). In one of the countries where the advocacy programme worked on strengthening diverse women's networks and organisations, they pursued the effective implementation of UNSCR 1325. Doing this supported and facilitated local partners in strengthening their role in the peace process and in pursuing women's rights and equality. The connection between global and local, however, was difficult, as the two realities were constructed around diverse languages. An advocate said, 'I have to translate our reports and label it into UN language' (advocate 2014, New York). The UNSCR 1325 framing of women's rights did not connect with the needs of the local partners. The programme manager stressed, 'I had been working [in this specific country] for a long time already, and on 1325 as well, but it was too sensitive […] to use 1325 language.' (programme manager 2014, The Hague). The local partners had their own framing of women's rights and their own national and local policies and laws in place. The programme manager stressed that 'there are a lot of differences; UNSCR 1325 is just a big label, but the issues are so specific in the different contexts' (17 February 2014, The Hague).

Adding to the issue of framing, the partners' strong sense of ownership also resulted in a less articulate or explicit advocacy role for the Dutch international NGO. The local partners developed and built women's networks, owning their progress and becoming more like independent partners than grantees. This made it difficult for the Dutch NGO to claim outcomes for these processes, as the Dutch programme manager emphasised:

> [This specific country] is difficult to have an impact linked to [our programme], because in [country X] the partners are very sensitive to ownership. This is about profiling, about putting a […] logo and stamp on processes, documents, etc. The link with [our programme] is there, but it is not visible per se. (programma manager 2013, The Hague)

A year after this comment was made, during a discussion with the programme staff, the advocate in the team stressed, 'We cannot afford to only be mentioned as a donor; we need to profile ourselves as thought leaders' (advocate 2014, The Hague). The lack of a visible link was a problem for the organisation in the Netherlands, as they needed to account for results by

claiming visibility. They needed to report on changes they had made possible. The Dutch policy officer emphasized this as follows:

> You have to put your name and logo everywhere; this is important for donors and results accountability […] It is about building a reputation and being seen as [a] thought leader. (policy officer 2014, New York)

The politics of visibility and the importance of profiling became clear during every event I observed. Organisational banners were set up, leaflets and reports from the organisation were stacked in piles on tables for people to take with them, tweets were sent out showing that the organisation was participating in the event and it was emphasised that events were organised with support from specific organisations.

All of this demonstrates that identifying advocacy outcomes was a political process at multiple levels, involving ownership over the processes as well as the need to account for outcomes in a certain way. The relation of the organisation's role and objectives with the local embedding was diffuse. Making outcomes visible involved diverse political interests. For the Dutch NGO, there was a vested interest in claiming the results achieved for accountability and for leverage reasons, to be seen as a thought leader in the specific field of women's rights. For its partners, the interest was in claiming ownership over their own processes in order to create leverage, while also translating and framing the language in their own political language.

Over-reporting

In other cases, outcomes were political in terms of over-reporting and claiming more results than were actually achieved. Outcomes and the reporting of outcomes in an evaluation assessment were a main priority for the funder (as indicated in the terms of reference and the evaluation questions highlighting outcomes) and for most of the programmes involved. Over the course of the evaluation, this resulted in many discussions over the meaning and assessment of outcomes (see Arensman and Van Wessel, 2017). These discussions provided insight into the political dimensions of outcomes in terms of accountability and legitimacy needs that sometimes resulted in outcomes being over-reported.

In a few cases, the evaluation team found that outcomes were portrayed as more powerful, influential and/or important than they were in reality (i.e. over-reporting) (see Arensman et al., 2017a). This was directly related to the significance ascribed to the evaluation results, which

could assumedly make or break a programme's future (i.e. in terms of funding or reputation). In one case, the following was claimed as an outcome:

> [The organisation] continues its conversation with the Gender Advisor of the Norwegian MFA during the New York advocacy week in October 2013, this time meeting with a group of Gender Focal Points to share developments on conflict prevention on their countries and provide feedback on Norway's involvement as an actor where relevant. (Internal outcome report, 2013)

During the advocacy week in October 2013, I was in New York to follow this advocacy process. The week was used by the organisation to facilitate its partners to come to New York. The organisation hosted sessions to bring its 'gender focal points' together to discuss the local and global updates, to make an action plan and to set an agenda for the coming year. There were 11 gender focal points from Fiji, the Democratic Republic of the Congo, Uganda, Burma, Serbia, Nepal, Canada, Lebanon and Azerbaijan, and the global secretariat coordinator from the Netherlands. The global secretariat of the organisation supported the members in terms of media contacts and training sessions, setting up meetings with relevant policy makers and providing members with the opportunity to contribute to the UNSCR 1325 week. The meeting reported as an outcome in the report quoted above was held during this week, and I was part of the meeting.

When I read the outcome statement, it surprised me that this meeting was chosen as an outcome for this advocacy week, because my notes from the meeting reflected a somewhat different situation. The meeting was located within the square mile around the UN compounds, and it was attended by the Norwegian representatives to the UN. The Norwegian government had previously been a donor to the organisation, and the meeting was set up as an 'update meeting, also to see whether the Norwegian government would be willing to contribute funding again', as the programme manager explained (2013, New York). However, the Norwegian representatives forgot about the meeting, and the group had to wait for half an hour before someone noticed us at the front door and let us into the building. When the representative was notified, she hurried towards the meeting room where we were waiting and stressed that she was very busy and had only 20 minutes left for the meeting. During the meeting, there was only time for a short introductory round of the network representatives; there was no time or space for critical discussions.

Whether this could be understood as an outcome is not entirely clear. The wording in the outcome statement appeared to overemphasise the importance of the meeting, but the meeting itself might have been important for keeping the network visible to possible donors and partners or for continuing to build the relationship with the Norwegian government. Consequently, again, while the outcome statement overemphasised the type of influence of this specific meeting, it underreported the meaning of the meeting for building and sustaining advocacy relations, generating visibility for the network and continuing steps towards progress on the gender agenda.

The politics of outcomes is directly related to the interests linked to the needs and demands for making outcomes visible. Outcomes become outcomes in relation to the meaning given to them by the advocates involved. The process of identifying outcomes is embedded in the interest in being accountable for the results achieved and the need to be seen as a credible and legitimate partner, advocate and grantee. This is about profiling as much as it is about the interests and contexts of the politics of results and the resulting pressure to make effectiveness visible. Working in the context of international development demands that outcomes be visible to demonstrate effectiveness but also to claim accountability and create legitimacy and credibility for the importance of the advocacy. There are interests involved in claiming or non-claiming outcomes, making them visible and being able to report on them. This makes the assessment and identification of outcomes, as well as the decisions around sharing specific information, political.

5.6. Discussion

Although research and practice on advocacy evaluation widely acknowledge the nonlinear, unpredictable and complex nature of advocacy, the problematic nature of advocacy outcomes remains understudied. Existing methods widely used for advocacy evaluation do not include reflection or tools for how to deal with outcomes that are not easily observed, made visible or measured. The findings presented in this chapter demonstrate that there is room for interpretation in the process of identifying advocacy outcomes, as the outcomes are constructed in multiple layers of diverse realities, arenas and stakeholders, and this involves a politics of outcomes.

Advocacy outcomes should be interpreted with the understanding that most of what is happening in pursuit of advocacy goals, strategies and outcomes is in the minds of the advocates involved. For evaluators, this means that much of the information necessary to identify

outcomes is in narratives as they are interpreted (i.e. given meaning) by the advocates. The advocates have their own understanding of what happened, how it happened, and what it led to, resulted in or spiralled into recursively. Meaning is thus given to outcomes in the interpretations of advocates' worldviews. Moreover, outcomes are constructed, as they come about in a multi-layered reality of multiple settings and arenas across borders, organisations and networks, involving multiple actors with diverse worldviews and multiple political and cultural realities. Meaning is given to advocacy in the construction of these multiplicities. The political character of outcomes is the negotiated character of what is being reported, shared or told and how it is given meaning in the narratives shared. The assessment and identification of outcomes is dependent on interpretations that are potentially be biased towards a more colourful picture of what happened, what the role of the advocates was and what was achieved. The 'honest story' therefore has to be dug out, questioned, verified and understood in terms of the meaning given to it through the interpreted, constructed and political process of outcome identification.

Evaluators and those being evaluated offer interpretations of events that are mostly not observable or visible in the form of straightforward changes. The relational character of advocacy outcomes makes them difficult to observe, follow or prove. Therefore, it is significant for evaluators to 'zoom in' on and dig out the construction of outcomes, including the room for interpretation and the politics involved. Current approaches like OM and OH do not consider these things to have bearing on the identification of outcomes. Although both OM and OH take a participatory approach to evaluation, they do not address the challenges to gathering and interpreting evidence or what these mean for evaluation or its findings (i.e. outcomes). These approaches are considered a means to an end (evaluating outcomes) and assume outcomes are logically measurable by following certain procedures. Both approaches analyse and interpret outcomes in relation to their significance for achieving the programme mission, goals or strategies (Smutylo, 2005; Wilson-Grau and Britt, 2012). The two approaches assume an objectivity to outcome evaluation that is based on the idea that outcomes are always visible and can be verified by independent sources of information; the approaches do not take into account demands and needs (claiming or non-claiming, over-reporting or underreporting), power inequality or dependencies (funding, access to information, resources), or how these things affect the evaluation and the findings.

This raised questions for me in my experience as an evaluator of advocacy processes. Neither OM nor OH considers the politics involved around outcome measurement and evaluation. They do not provide guidelines or directions for working with outcomes that are not specific, visible, measurable, observable or objectively identifiable. Finally, these approaches overemphasise

measurable outcomes, at the risk of missing out on the meaning of how outcomes come about, including overlooking those processes that do not necessarily lead to (visible) outcomes.

My findings demonstrate that a participatory approach is indeed necessary to evaluate advocacy. This is not only because evaluators need access to information about theorising and strategizing, which is in the heads of advocates, but also to stimulate learning together with the advocates involved. My insight into strategic non-claiming would not have come out without meaningful and open interactions with those involved. Additionally, the politically sensitive nature of making outcomes visible, claiming those outcomes, and, thus, over-reporting and underreporting would have gone unnoticed. However, this kind of participatory approach presents challenges with regard to the identification of outcomes that are not self-evident. As my findings show, advocacy outcomes are given meaning in narratives, which are interpreted by those involved and constructed in multi-layered realities. These narratives are also political, because much of the legitimacy and credibility of advocacy is dependent on the ability of the advocates to demonstrate effectiveness. The current approaches do not provide solutions or tools to deal with these kinds of challenges.

My findings call into question the ideas that independently gathered information exists and that an objective evaluation is necessary to make a valuable assessment. The findings demonstrate that advocacy outcomes are very difficult to measure objectively, which means that advocacy effectiveness is likely not objectively measurable. Identifying outcomes and judging effectiveness are often more related to the meaning given to outcomes in terms of the legitimacy and credibility of the organisation than to some kind of objective measurement. For this reason, I argue that there is a need for a different understanding of outcomes, shifting from the narrow outcome-driven focus on measurable results to an understanding of outcomes as given meaning through political interactions and constructions that provide room for interpretation. Outcomes are much more than just the result of achieved objectives; they can also provide insight into the politics and meaning attributed to results in certain contexts.

For evaluation, this means that evaluators and the methods they use must improve in terms of their ability to identify meaningful outcomes. This demands changing the way we think about outcomes. Outcomes need to be understood in relation to the processes of advocacy as strategic practice, including the idea that outcomes are pursued and strategized for in ways that do not always follow a predefined path. This kind of understanding means 'zooming in' on the dynamics that give meaning to outcome processes and practices, including how they fail to come about. Identifying outcomes and their challenges requires the flexibility to incorporate an understanding of the roles played by interpretation, construction and politics. This is especially

the case for advocacy because there is a dependency on the advocate's practical judgements and account of events to understand the processes and practices leading to results—or the lack thereof.

Evaluators and practitioners need to acknowledge that the systems shaping evaluation standards are formed by how effectiveness is defined and that outcomes have been given centre stage in terms of effectiveness claims. Evaluating effectiveness requires a clear understanding of what is being evaluated, why and how, as well as what evaluation questions need to be answered to assess a certain kind of effectiveness. Effectiveness is not only established in the identification of outcomes, but also in the interactions between the processes of theorising and strategizing, with practices that demand acting, reacting and adapting to changing circumstances.

5.7. Conclusions

The findings provide important insight into the problematic nature of outcome identification, which is especially relevant for the field of advocacy evaluation. Working in an outcomes-focused manner together with the pressure to demonstrate effectiveness through measurable results (Best, 2017; Eyben et al., 2015; Riddell, 2014; Vähämäki et al, 2011) risks adopting evaluation approaches that trade quality for quantity. These kinds of approaches may overemphasise tangible outcomes in light of what was planned at the expense of understanding the process of how these came about. Advocacy evaluation as a field of practice is struggling to come to terms with the pressure and need to demonstrate effectiveness to account for how funding is spent and to attain legitimacy and credibility for the organisation or programme, while also coping with the problematic issue of measuring advocacy effectiveness and questions of how to assess effectiveness of processes that are complex, dynamic and difficult to measure.

Although outcomes are increasingly important in the pursuit of effectiveness, they are still considered self-evident and are therefore understudied. The main point made in this chapter is that advocacy outcomes are not self-evident, but rather political and constructed through multiple realities, leaving much room for interpretation. This influences not only how outcomes are interpreted (e.g. claiming, non-claiming, over-reporting and underreporting), but also the quest for specific types of outcomes (e.g. those that are easily measured, made visible and claimed). Identifying outcomes in advance may lead to overemphasising certain outcomes over others, as well as 'sticking to the plan' rather than revisiting the strategy and its outcomes. This

leads to the another point: In advocacy, adaptation, strategic interactions and decision making, and reflection are crucial.

If outcomes are not self-evident and we need to identify outcomes that are meaningful to the processes and practices of advocacy, this demands more effort from evaluators to surface, question and critically reflect on the meaning of effectiveness. Understanding advocacy outcomes is a quest. The processes need to be (re-)constructed, questioned and interpreted in an inquisitive way that places advocates' strategic action at the centre, instead of merely focusing on what is achieved. This also demands that evaluators seek out experiences of what advocacy did, how and why: Why were certain directions taken? What were the underlying considerations? What made the advocates act or interact in that way? Being inquisitive by focusing on the experiences and the stories that give meaning to everyday practices requires an evaluator to have an open mind and to take an approach that is truly interaction- and people-centred by focusing on the advocates, their roles and understandings, the diverse realities in which they operate, and their strategic and practical judgements, considerations and actions. In this way, Coffman's (2007) interim changes would also acquire a more meaningful place in the identification of outcomes as a strategic process.

This chapter has provided insight into the problematic identification and evaluation of advocacy outcomes. To continue building on the findings presented above, I encourage other researchers to develop what we now know about advocacy effectiveness and evaluation into concrete steps and approaches guiding evaluators and practitioners to really do justice to the dynamic nature of advocacy in evaluation.

6. Putting the puzzle together: Conclusions and Contributions

6.1. Introduction to the conclusion

Although advocacy is increasingly implemented as a strategy for NGO effectiveness, it challenges existing evaluation methods that often seek to measure and predict effectiveness in terms of quantifiable results (Devlin-Foltz et al., 2012). With the ever-growing pressure around effectiveness, the public sector compelled to make their results visible (Best, 2017; Eyben et al., 2015; Kremer et al., 2009; Riddell, 2014). This affects protocols and procedures, as well as how evaluations in different traditions deal with the measurement of effectiveness. Key discussions on evaluation and evaluation methods are based on the idea that evaluation is objective, seeking general patterns of cause–effect relations to predict and control effectiveness (Armytage, 2011; Marchal et al., 2012; Riddell, 2014). This 'objective' approach to evaluation assumes an independent relation between the evaluator and the evaluated and demands the replicability of results (Guljarani, 2011; Mansoor, 2003; Van der Knaap, 2004). Advocacy effectiveness is not easily measured under the terms and conditions driven by the political pressure for measurable, visible and tangible results. What, then, constitutes advocacy effectiveness, especially regarding the dynamic character of advocacy and of the contexts in which advocates and evaluators operate?

The aim of this doctoral dissertation was to contribute to understanding how the practice of advocacy for development gives meaning to effectiveness and how the politics of results play out in the evaluation of advocacy. I define effectiveness as the processes and practices through which an intervention reaches its goals, which is less about whether the intervention does reach its goals and more about how the intervention reaches these goals. The different chapters in this dissertation demonstrate that whether and how a course of action was effective are not self-evident facts and that defining the meaning of effectiveness is an intrinsically interactive and political process. Because outcomes and effectiveness are difficult to establish in advocacy, the question of how outcomes are reached is important. The process of defining outcomes and effectiveness is profoundly political; I call this process the politics of results. More specifically, the politics of results is the negotiated process by which outcomes and effectiveness are given meaning in the practice of advocacy for development. Because the outcomes and effectiveness of advocacy for development are difficult to operationalise, in practice, it turns out that the

politics of results often boils down to the question of whether the intervention or organisation succeeded in achieving legitimacy and credibility in the eyes of other stakeholders.

Over the last decade, advocacy evaluation has grown into a field of study in its own right, with a growing body of literature on advocacy effectiveness and its evaluation (see Beer and Reed, 2009; Coe and Majot, 2013; Coffman, 2007, 2009; Gardner and Brindis, 2017; Harvard Family Research Project, 2008). Existing studies provide important insight into advocacy as a complex practice and discuss methods for advocacy evaluation (see Table 2 in Chapter 1). However, the existing literature has not discussed certain important dynamics, such as how effectiveness is given meaning, the role of politics, the problematic nature of evidence in establishing effectiveness and the challenges associated with measuring effectiveness. The latest extensive study of advocacy evaluation, conducted by Gardner and Brindis (2017), provides an overview of advocacy evaluation practices and theories. However, their book does not provide tangible guidelines for how to deal with nonlinearity, how to understand advocacy as a messy process, how to deal with the struggle for evidence or how to account for the fact that outcomes are not self-evident. I have tackled these issues in this dissertation and explored what advocacy processes and practices look like, what this means for effectiveness in terms of outcomes, the relation between theories and practices, and the implications of all of this for advocacy evaluation.

The findings presented in this dissertation contribute to a better understanding of the relational aspects that give meaning to effectiveness. The previous chapters have all related to answering the central research question: *How does the practice of advocacy for development give meaning to effectiveness, and how do the politics of results play out in the evaluation of advocacy?* Answering this question is approached by unpacking how effectiveness is pursued, negotiated and interpreted. Section 6.2 below summarises the main findings of the empirical chapters and answers the research questions outlined in Chapter 1. Next, Section 6.3 discusses the cross-cutting findings and theoretical implications. I contributed to theory strengthening in two main ways, relating to the politics of results and the meaning of advocacy effectiveness. Section 6.4 then discusses the practical implications of this work for advocates, evaluators, policy makers and others working with or in advocacy or evaluation. This section also points to potentially fruitful areas for further research to explore, based on the implications of the present study's findings.

6.2. Main findings: Conclusions from the four empirical chapters

When starting this research in 2012, I felt that a lot of politics were at play in evaluation, because politics are an inherent part of human interactions—including interactions about conceptualising effectiveness; interactions focused on identifying, assessing or presenting outcomes; and strategic interactions and negotiations that take place as part of evaluation. Indeed, especially around results, there were many interest-driven processes regarding what information was accessed and how the results were identified and presented. Being a researcher as part of a team conducting a major evaluation provided the unique opportunity to observe, follow and study these processes. Along the way, I learned that the 'honest story' about effectiveness is much more layered and multidimensional than I had thought.

Existing studies examining the politics of evidence and results (Best, 2017; Riddell, 2014; Eyben et al., 2015; Natsios, 2010) have tended to focus primarily on the pressure for results in international development, as translated in the power inequalities among donors, recipients of funding and beneficiaries of aid interventions. Some of these studies (e.g. Riddell, 2014) have stressed the consequences of political pressure for results, noting that this leads to a narrow focus on measurable, visible and short-term outcomes and to overlooking the actual development process in international development interventions. A few studies have discussed what results-pressured politics means in terms of procedures and protocols in the development world (Best, 2017; Guyt, 2015). However, these studies have not generally touched upon the meaning of the politics of results in the identification, measurement, presentation and interpretation of effectiveness (i.e. outcomes), or how this affects evaluation. There are no studies that specifically discuss advocacy or advocacy evaluation in relation to the politics of results.

In line with the existing trends of the politics of results in results-based management and in positivist traditions of evaluation, effectiveness is often not understood from a perspective of everyday practices and politics. Rather, effectiveness is analysed from a perspective that focuses on political interests and dominant discourses that interpret effectiveness in terms of measurable, tangible and short-term results. This approach is in tension with the complex reality of advocacy, as it does not provide space for analysing how the meaning of effectiveness comes about through the interplay of factors that make up the practice of advocacy—factors that are not necessarily measurable, predictable, controllable or tangible. My findings contribute to this discussion in multiple ways, providing empirical illustrations of the politics of results beyond donor–recipient relations (power and resource distribution). The analysis discusses navigating, negotiating and manoeuvring around interests, needs and demands in evaluation. This demonstrates that effectiveness is given meaning not only in terms of results, but also through

everyday practices, involving the politics of results. Evaluating advocacy effectiveness requires more than reviewing the achievement of stated objectives; it calls for unravelling the interplay of factors through which effectiveness is shaped: its pursuit, negotiation, diverse understandings and representation.

The individual empirical chapters of this dissertation present an analysis of the three advocacy programmes I was closely involved in evaluating, and, as a fourth case study, the MFS II ILA evaluation[32] itself. These chapters demonstrate how the practices of advocacy for development give meaning to effectiveness and how the politics of results play out in the evaluation of advocacy. Chapter 2 treats the multidimensional nature of advocacy effectiveness by focusing on a case study of a transnational advocacy network (TAN) in which ownership and effectiveness were simultaneously pursued. The case study presented in Chapter 3 is the evaluation itself, where effectiveness was negotiated in the identification, measurement and presentation of outcomes. Chapter 4 provides an analysis of an advocacy programme on child rights. Here, advocacy is shown to be a strategic and recursive process of practices that happens at the interface between theory and practice. The chapter proposes a 'twinning' of theory of change (ToC) with practices of change. Finally, Chapter 5 includes all of the case studies, showing how advocacy outcomes are not self-evident but rather interpreted, constructed and political in nature. The findings of the four chapters are summarised below, after which I discuss the main overarching findings of the research.

6.2.1. Advocacy effectiveness is multidimensional

Ownership is considered an important element in effectiveness discussions in the context of international development. OECD DAC (2005, 2008) stresses the need for country ownership in pursuing effectiveness of development aid, and numerous civil society organisations including NGOs strive for ownership in their pursuit of effectiveness (Faust, 2010; Fowler, 2000; Reich, 2006; Richmond, 2012). Although international development policies and practices consider ownership a condition for achieving effectiveness, the implementation and pursuit of ownership and effectiveness involve different practices and understandings. Effectiveness is often treated as an instrumental concept, with results seen as unambiguous and rational. Ownership, in contrast, is a socially constructed reality that acknowledges the importance of diverse perspectives. To make sense of how these concepts operate in practice,

[32] Abbreviated from: Netherlands Ministry of Foreign Affairs Multi-Financing Scheme 2010-2015 International Lobbying and Advocacy Evaluation.

Chapter 2 focused on a TAN that pursued ownership and effectiveness simultaneously, exploring the following question:

To what extent are ownership and effectiveness pursued by a TAN in tension with each other? What does this mean for the advocacy effectiveness?

Ownership and effectiveness are both global policy demands, and both are often included in donor–recipient agreements as part of the funding relations. However, the two concepts continually challenge and define one another. Effectiveness is demanded in measurable, presentable terms, whereas ownership does not readily lend itself to measurement. Moreover, ownership demands a different understanding of effectiveness—as an ambiguous, multidimensional concept that acquires meaning through interactions between social and political considerations. The existing literature focusing on effectiveness in TANs (Saz-Carranza, 2012; Wong, 2012) does not call into question the nature of advocacy effectiveness or the relation between the pursuit of effectiveness and the simultaneous pursuit of ownership. In the present research, I found that a strong emphasis on ownership may generate a different understanding of effectiveness—one in which the formation, operation and linkages of a network can be seen as both means and results of advocacy work. The politics of results in international development frames ownership as being in direct relation with effectiveness. This assumption is questioned in Chapter 2, as the findings demonstrate that the relation between the two concepts is ambiguous.

Although the studied TAN found a discursive solution to the contradiction between ownership and donor-defined effectiveness by bringing process outcomes into its definition, the pursuit of ownership and the simultaneous pursuit of effectiveness continued to challenge and define one another in practice. The study presented in this chapter demonstrates that ownership is nurtured in the inclusiveness of diverse realities sometimes leading to partial collectiveness, whereas the (genuine) drive to realise outcomes in an effective way requires a balance between unity and diversity to ensure a strong collective.

The findings also show that effectiveness is often negotiated amongst network members. However, effectiveness is considered and pursued in a technocratic way, searching for collective and measurable changes, which is in contradiction with the aim of inclusiveness. Especially when the pressure to report outcomes increased, the network approach of inclusiveness did not fit the technocratic ideas of effectiveness. The network also felt increasingly pressured towards a technocratic approach to effectiveness in terms of producing

measurable results, which clashed with the network's desire to be inclusive. At the same time, the advocates were also intrinsically motivated and driven to pursue structural changes to make the world a better place. They had a vested interest in showing effectiveness to create legitimacy and credibility for accountability reasons, as well as for building their reputation and cause. This brought out how, under pressure, tensions around commitment and partial collectiveness resulted from the dynamics between the pursuit of ownership and the demands for technocratic, measurable effectiveness. This resulted in a power game that turned effectiveness in a technocratic straightjacket, rather than making it part of an inclusively defined set of goals for advocacy.

6.2.2. Advocacy effectiveness is negotiated

Based on an analysis of the MFS II ILA evaluation as a process (focusing on the three specific case studies), Chapter 3 argues that advocacy evaluation is necessarily subjective. This subjectivity provides for interesting opportunities to probe, reflect on and develop insight into the value of advocacy work. Building on this analysis, this chapter answered the following question:

How is effectiveness negotiated, and how do interactions in advocacy evaluation give meaning to advocacy effectiveness? What implications do the negotiations have on evaluation?

The politics of results affect evaluation theory and practice by putting outcomes at the centre of effectiveness and by assuming these can be objectively identified, measured and presented. The literature includes only a limited discussion of how this influences what is evaluated, how it is evaluated and why is evaluated. Chapter 3 argues that a shift or reconsideration is necessary in the current positivist trends (i.e. performance measurement, results-based management), especially because advocacy effectiveness is not measurable in the rigorous manner that these kinds of methods pursue (i.e. tangible, visible, short-term outcomes).

The findings presented in Chapter 3 demonstrate how effectiveness was given meaning through a collective and iterative process in which the identification of outcomes was co-constructed through dialogue and investigation, and the measurement and presentation of outcomes was discussed and negotiated between advocates and evaluators. These interactions were valuable in making room for differences, multiple interpretations and diverging understandings. However, this also led to confrontations about the extent to which space should

be given to do justice to the complexity and challenges of advocacy and about where it could be concluded that more effectiveness would have been possible.

Besides clarifying the co-constructed and negotiated process of giving meaning to advocacy effectiveness, this chapter offers a way to make room and generate understanding of the diverse realities in evaluation. This includes the vested interests involved for evaluators seeking to make a critical assessment and for those being evaluated, who will not always find these assessments to be exactly what they want to hear. Taking a constructionist approach can support strengthening evaluation accuracy and quality by giving space to understanding these diverse realities, adjusting preconceived notions of outcomes and measurement, considering and investigating diverse levels of evidence, and providing room for a better understanding of the challenges faced by both evaluators and advocates. The highly political nature of evaluation is often unrevealed or misrepresented with the current methods and approaches to evaluation, even though it has bearing on the evaluation process and outcomes.

6.2.3. Theory of change needs to be twinned with practices of change

ToC is *hot* in evaluation practice, as well as in the development world, where it is being used and implemented by donors, policy makers and NGOs. A ToC approach is considered especially suitable for complex and dynamic processes. In Chapter 4, I question whether the ToC approach lives up to this expectation of suitability for complex and dynamic interventions such as advocacy. ToC began as a theory questioning how and why an intervention works (Weiss, 1997, 2000). Currently, ToC is used to acquire funding (explaining how complex interventions plan and achieve outcomes), as it is seen as a roadmap to plan on getting from 'here to there' (Stachiowak, 2013; Stein and Valters, 2012; Valters, 2014). In evaluation, it is used to assess the relation between planning and achievements, questioning the underlying assumptions regarding the steps and strategies undertaken (Beer and Reed, 2009; Vogel, 2012). In Chapter 4, I focus on important limitations of ToC as an approach to evaluating advocacy, arguing that, although ToC builds in more flexibility, its focus remains on cause–effect relations and intended outcomes, which does not leave much room for advocacy dynamics. Chapter 4 answered the following question:

What is the relation between advocacy processes and practices and the theory-driven methods used for advocacy evaluation? Does a ToC approach to advocacy evaluation do justice to advocacy processes?

ToC as an evaluation approach is valuable and an important contribution to new trends in evaluation that provide space for reflection and the critical questioning of assumptions underlying change processes and practices. However, I argue that, as a method for advocacy evaluation, ToC remains too stuck in theory, while the recursive, unpredictable and nonlinear nature of advocacy demands more space to account for practices of change. ToC as it is currently used does not live up to its potential to guide complex processes such as advocacy. It does not provide the necessary tools for evaluators to understand the interaction between theory and practice. Understanding dynamic processes such as advocacy demands 'zooming in' to examine the theory on which strategies and assumptions are based in relation to practices—how advocates act, adapt, manoeuvre and strategize in their everyday realities. Existing work has not necessarily provided a tangible, overarching understanding of what this means for advocacy and its evaluation. This chapter demonstrates that change is initiated from strategies as practices and is shaped through human interaction.

Current evaluation methods are limited in terms of doing justice to the dynamic nature of advocacy in the evaluation of its effectiveness. Scholars have emphasised the need to look at the advocates instead of the outcomes, because advocacy is inherently political, making it unpredictable and nonlinear (Jones, 2011; Teles and Schmitt, 2011). However, these studies have not fully explained what this means for advocacy evaluation practices, and the currently widely used methods, such as ToC, do not put this to the test. They do not provide evaluators with tools for how to better grasp the meaning of practices or how to put advocates at the centre of evaluation. Chapter 4 provides an analysis of the strategic practices of advocacy as recursive and of the advocates as relating to this recursiveness strategically, leading to emergent rather than linear or predefined outcomes. Part of the aim here was to provide a starting point for evaluating advocacy by twinning theory with practices.

Twinning ToC with practices of change opens the door to new ideas, providing space to put advocates at the centre of evaluation, because their practical judgements and manoeuvring of situations and stakeholders shape their interpretations of events that give meaning to effectiveness. Twinning ToC with practices of change bridges a connection between the theory of evaluation methods and strategy as practice—something actors *do* rather than something organisations *have* (i.e. strategy-as-practice and recursiveness, as seen in organisational studies). Analysing these strategic practices of advocates is key in the effort to put advocates at the centre of evaluation. This requires an understanding of advocacy as strategic and recursive, which, again, leads to emergent rather than linear or predefined outcomes. Effectiveness is

given meaning through the strategic considerations and decision making that inform and inspire advocates as drivers of change. This gives room to understanding everyday practices, while also relating these practices to the (changing) theories (plans, strategies, objectives).

6.2.4. Advocacy effectiveness is not self-evident

In establishing effectiveness, outcomes are increasingly important. In the tradition of New Public Management, effectiveness and performance in terms of outcomes achieved by the organisation are a main focus for evaluations (Best, 2017; Brinkerhof, 2008). These results-driven methods focus on making improvements that allow organisations to achieve better results (Hatton and Schroeder, 2007; Hulme, 2010). In this tradition, evaluations are outcome-focused, building on the assumption that outcomes follow self-evidently from action. However, with advocacy evaluation, the outcomes are not self-evident. Identifying advocacy outcomes is difficult because the processes are often intangible and impossible to observe or trace, as they arise from personal interactions. Chapter 5 therefore aimed to answer the following question:

How are advocacy outcomes identified in advocacy evaluation, and what are the challenges of and lessons learned from this process?

In Chapter 5, I demonstrate that advocacy outcomes are not self-evident. The identification of outcomes is dependent on the diverse interpretations of the stakeholders involved. Outcomes are constructed in multiple layers and are built across borders. This process involves multiple actors, as well as networks that often include advocacy targets. Outcomes are political in terms of how certain outcomes are claimed or reported while others are specifically not claimed. In some cases, outcomes are identified (claimed, not claimed, over-reported or underreported) to fit the purpose of the evaluation (i.e. showing measurable results). In other cases, identifying outcomes and judging effectiveness is more related to the interpreted (diverse realities), constructed (multi-layered) and political (multidimensional, multiple interests involved) character of outcomes than to some kind of 'objective' measurement.

The findings presented in this chapter build an argument for changing how evaluators and practitioners think about outcomes in evaluating advocacy. Most of the information necessary to identify and assess advocacy outcomes is in the heads of the advocates involved, as they act and interact with stakeholders while adapting to changing circumstances. This makes the identification of advocacy outcomes a process in which politics and interests play a role. The

interpretation of narratives also becomes important for surfacing how outcomes came about, who was involved, what happened and why it happened. This makes it extremely important for evaluators to talk to different people involved in the process, including advocates, evaluators, other staff members, partners and advocacy targets.

Advocacy outcomes need to be (re-)constructed, questioned and interpreted in an inquisitive way, and evaluators need to pay more attention to the processes and practices through which advocacy effectiveness is given meaning and to the way outcomes are identified. This requires evaluators to keep an open mind and to genuinely take an interaction- and people-centred approach by focusing on advocates, their roles and understandings, the diverse realities in which they operate, their strategic and practical judgement, and their considerations and actions. Inquisitive evaluation must focus on the advocates and their practical and strategic actions, rather than on the outcomes. Evaluators have to seek out and question narratives, practices and underlying assumptions.

The findings presented in Chapter 5 provide insight into the quest of evaluators to identify advocacy outcomes, demonstrating that advocacy effectiveness is given meaning through the interpreted, constructed and political character of this process. Questions of *What outcomes?* and *Whose outcomes?* matter as much as questioning how the outcomes came about, and effectiveness cannot be established by focusing merely on what was achieved. Existing approaches provide important tools to build on, but these approaches are limited when it comes to outcome identification. Although methods like outcome mapping and outcome harvesting do take a participatory approach, working closely with those involved (i.e. programme managers, advocates, evaluators), they do not question the nature of outcomes and their identification. In identifying outcomes, it was especially challenging for the evaluation team observed in the present study to reconstruct outcomes and how they came about. Moreover, because of the interpreted, constructed and political character of advocacy outcomes, evaluating advocacy effectiveness is necessarily subjective. This raises critical questions around evidence (e.g. When is what kind of evidence good enough for a judgement call?), challenging the idea of an objective evaluation that focuses on outcomes as self-evident and measurable—an idea subscribed to by both outcome mapping and outcome harvesting. This poses challenges for evidence gathering that have not been discussed in existing methods.

6.3. Main findings: Cross-cutting conclusions

The literature and common practice in advocacy evaluation consider effectiveness and the (measurement of) outcomes to be self-evident, but the findings of the empirical chapters of this dissertation demonstrate the opposite to be true. The ideas that outcomes are self-evident and that effectiveness is measurable shape the protocols and procedures around measuring effectiveness. This leads to critical limitations with regard to identifying outcomes, meaning-making and negotiations. Although the existing advocacy evaluation literature and common practice in the field acknowledge the complex and nonlinear character of advocacy, this acknowledgement does not translate this into a consideration of what this means for advocacy effectiveness or the problematic identification of outcomes (see Table 2 in Chapter 1). For the theoretical embedding of the findings, I found more direct value in literature from sociology and organisational studies about performance, strategies-as-practice, communication and recursiveness (Crozier, Guba and Lincoln, 1989; Herman and Renz, 1999; Jarzabkowski, 2005); ideas about politics as a human- and interaction-driven process (Arendt, 1998; Mouffe, 2005); and the importance of everyday politics, effectiveness and the politics of evidence and results in development studies and anthropology (Eyben et al., 2015; Hulme, 2010; Kerkvliet, 2002). A first cross-cutting finding is therefore that, especially given the transnational, political and dynamic nature of advocacy for development, more cross-fertilisation with the diverse bodies of knowledge (the literature referred to above) is necessary. The final question addressed here relates to all of the empirical chapters in this dissertation:

What do the findings in the previous chapters mean for advocacy evaluation?

In answering this overarching question, I identified two cross-cutting findings.

6.3.1. Formal and informal meaning-making

Throughout the evaluation, there was a palpable tension between formal and informal evaluation procedures. Informally, discussions and dialogue took place between the evaluators and the evaluated. In these discussions, outcomes were surfaced in co-creation, and challenges around identifying and measuring outcomes were discussed. This helped us as evaluators to establish how and why certain strategies were implemented, and to what end. In other cases, this dialogue helped in gaining insight into the recursive nature of advocacy or understanding

about how the advocacy work led to emergent outcomes that were not intended or pursued from the start. This informal dialogue revealed the fundamental dynamics, challenges and contexts in which the considerations, negotiations and interpretations underlying strategic decision making were formed. Through this process, outcomes that were not reported or claimed were found, and it was also possible to see how certain outcomes were over-reported or claimed although it was questionable whether these were actually outcomes. These informal discussions created a space of openness and learning.

Formally, the evaluated organisations were keen to present their outcomes (results) as more successful (over-reporting, claiming visibility), especially in writing. There was an evaluation report to be written, and it was the evaluators who had the final say in terms of what would be included and how it would be presented in the findings. As soon as the findings were written down, the relations between the evaluators and the evaluated came under strain. Organisations pushed for a more positive presentation of the findings or for certain less successful, challenging or sensitive elements to be left out. This resulted in pressure on the evaluation team to adapt the findings. This tension between the informal and formal relations presented itself as a tension between theory and practice. In practice, organisations said that they did one thing (e.g. pursuing networking, cooperation and influencing) while, based on theory, they reported on another level (measurable indicators).

Meaning-making around effectiveness thus happened in informal discussions and dialogue, as well as in formal negotiations on what constituted results, how they came about, what challenges were experienced and how to deal with those, and how the successes and failures should be presented. In these processes, the space for interpretation resulted in multiplicities of meaning of effectiveness, with people in different positions understanding effectiveness differently. For network members in the studied TAN, effectiveness could mean the internal processes of networking, network strengthening and knowledge sharing. For the network secretariat, effectiveness could be forming an inclusive collective message even if it was not strongly embedded in a strategic campaign. For another organisation, effectiveness was the many events organised and policy papers produced, even though the influence of these outputs was not really observable and was thus not included as part of their understanding of effectiveness. For others, it was gaining visibility as a 'thought leader' that counted for effectiveness, because this resulted in being invited to decision-making tables. In all of these cases, informally, stakeholders were open to discussing all kinds of challenges, but, in their view, these were not meant to be included in any reported findings.

This difference between formal and informal meaning-making creates an image of two different worlds in which practitioners and evaluators work. There is one world of formal accountability structures, where power is divided along the lines of resource distribution. In this world, procedures and systems are in place that demand successful, visible changes as results. This may require overemphasising the role or influence of the advocacy programme. The other world is that of messy practices, where results may not be straightforward or even visible and knowledge is tacit. These complexities require reflection and introspection about roles, changes and how changes are theorised and perceived, while also acknowledging that challenges occur and that changes and roles in pursuing changes are not always clear.

6.3.2. The politics of results is multidimensional

The practices of advocacy and the evaluation of its effectiveness are intertwined with the multidimensional reality of the politics of results. The politics of results is part of the way advocates challenge, contest or comply with norms, values, rules and power structures in their pursuit of effectiveness. If one were to look for politics only in conventional places and forms, much would be missed in terms of political thought and action, as well as the relationships between political life in strategic decision making and the political systems where these relationships are found. The theoretical and practical complexity of advocacy for development is in tension with the manner in which effectiveness is promoted and understood in the broader development world as necessarily measurable and objective, and this tension is part of the politics of results. These two realities exist simultaneously, reinforcing and challenging one another. Models and theories that reinforce the predictability of effectiveness in measurable results have been developed to make sense of the complexity of development processes. However, this process obstructs the understanding of effectiveness in more complex terms. In this process, effectiveness is given meaning through the dominant structures in the development world as being measurable, predictable, tangible and visible within the time frame of funding cycles. At the same time, the meaning of outcomes is contested, negotiated and challenged in the diverse realities in which advocates and evaluators operate and cooperate. Understanding the politics of results provides room for these negotiations and interpretations of results.

The politics of results is multidimensional in character. It is about pursuing, negotiating, understanding and identifying effectiveness. This politics appears in the form of interests that manifest themselves in struggles between the needs (legitimacy, credibility) and demands (accountability) of both the evaluators and the evaluated. The formal procedures and processes

in development are directed towards accountability and the objective evaluation of measurable results but, again, the case studies in this dissertation reveal that the meaning of results, and thus of effectiveness, is negotiated and contested, as it is embedded in diverse realities.

One of the findings presented in this dissertation is that advocacy is necessarily subjective, challenging conventional evaluation methods that seek to provide an objective perspective on effectiveness. Accounts of what happens, why and how are important for understanding effectiveness, but these are also part of the politics of results through which organisations seek to portray their achievements in a particular way (with certain results, indicators or successfulness). The politics of results is about how to portray and interpret advocacy actions and achievements, as well as the meaning attached to these achievements in relation to the needs within the organisation as well as vis-à-vis other stakeholders and the donor. Here, outcomes are not objectively identified, and effectiveness is not objectively measurable and cannot be separated from the politics around results. Identifying advocacy outcomes and judging advocacy effectiveness in this context are mostly based on the interpretation of events by the advocates and on the credibility and legitimacy of the organisation.

6.4. Main findings: Practical relevance and directions for future research

My findings have implications for practitioners such as advocates, evaluators, funders and policy makers. The findings also have a wider relevance, beyond just those working directly with advocacy, because they provide insight into everyday practices and processes that are shaped by interactions and are not easily planned, measured or made visible. These findings can thus also be applied to other dynamic and complex processes in which effectiveness is pursued and evaluated, such as most human-driven processes—for example, development interventions or organisational effectiveness studies.

6.4.1. Advocacy evaluation demands a different approach

What implications do these findings have for evaluators and practitioners in advocacy for development? The most important implication is that advocacy evaluation demands a different approach to evaluation. Current evaluation methods do not do enough justice to or support and accommodate the dynamic processes of advocacy. None of these methods (i.e. outcome mapping, outcome harvesting, contribution analysis, ToC) was developed specifically for advocacy, but they are nonetheless widely used in the field. Existing methods following the

ToC approach (Vogel, 2012), while valuable in providing the right directions and input for encouraging more flexible and reflective ways of working, have not yet lived up to their potential because there is a gap in terms of actually understanding and dealing with such complex practices. Methods such as outcome harvesting and outcome mapping (Simon and Hearn, 2009; Wilson-Grau and Britt, 2012) also provide important elements for mapping, strategizing and surfacing outcomes; however, in evaluating dynamic programmes like advocacy, they tend to overlook the important 'how' and 'why' questions that give meaning to the processes and practices. Many of these methods focus solely on outcomes, with an if–then or cause–effect logic (i.e. *if* I do this, *then* it will have this result). This does not provide the space or tools to convey the meaning of the outcomes or the processes by which these meanings came about.

Advocacy evaluation demands an approach that focuses on the practices and practitioners. Evaluators and advocates need to understand and act on the fact that knowledge about advocacy effectiveness and practices is often tacit. I have argued that advocacy processes and practices are not only recursive and emergent in nature; they are, for the most part, also interaction-driven. Much of what happens is strategically acted upon by advocates, who move in changing environments; negotiate and manoeuvre through interests, needs and demands; and interact with other stakeholders. Consequently, in the field of advocacy, there is no agreed view on 'what works when'. Although there are some shared understandings as to how to act effectively (e.g. based on an analysis of the political context, timeliness, engaging with policymakers' needs and understandings, providing credible evidence and usable ideas, and working in coalitions), these understandings do not explain what is sufficient to achieve success. Advocates strategize based on practical judgement and opportunities for change. Much of the information about how things happened, what happened and with what consequences or results, and why things happened as they did are inside the advocates' minds and understandings of the state of events. Because of this, again, much knowledge about advocacy and its effectiveness is tacit in nature. This knowledge exists and develops in advocates' heads and in interactions. Understanding and acknowledging this can also help in making sense of other complex and dynamic processes where tacit knowledge is key. Doing justice to advocacy therefore requires a different evaluation approach that puts the advocates, interactions, strategy-making and considerations at the centre.

The findings presented in this dissertation imply that evaluations need to refocus, shifting away from measuring outcomes to make human interaction, the agency of stakeholders and strategic practices the starting point for understanding effectiveness. This means a shift away

from narrowly focusing on effectiveness as predictable and controllable, and on outcomes as measurable and visible. This requires evaluators to move away from the cause–effect thinking of *if–then* logic towards taking an actor perspective and a *how-did-it-evolve* logic (Chapter 4). Evaluation, then, should be much more open, inquisitive and qualitative in its approach to data collection. This demands that evaluators critically question and investigate advocates' strategic decisions and narratives around how change happened (Chapters 2, 4 and 5), asking how did what kind of interactions lead to certain kinds of strategic decisions. This involves questioning the interpretations and validating these with evidence from diverse sources such as partners, other civil society organisations, advocacy targets, written documentation and records of interactions (i.e. emails).

The politics of results implies that evaluators also have to take into account their own roles and assumptions, the role and meaning of the evaluation questions (what is that you want to know, and for what purpose?), and the effect of the dominant paradigms in the context in which the evaluators and the evaluated work. An important part of this point involves the evaluator's own role in the space for the interpretation of results (Chapters 3 and 5). Meaning that close cooperation between evaluators and programme staff, advocates and targets is necessary and that trust-building and transparency (i.e. open discussion, learning, joint reflection) are very important. Depending on the purpose of the evaluation, learning and co-construction should be included in the interactions between the evaluators and the evaluated. This provides room for understanding the different realities, the strategic navigations of advocates, how decisions are being made and what results from these decisions.

Notably, and with clear justification, organisations and donors alike will want to know whether their money has been well spent. The current focus for this point is on the utility of effectiveness (accountability, legitimacy, credibility), which is wrapped up with the performance of organisations (i.e. the technocratic perspective of measuring visible changes) and the assumption that effectiveness is measurable. This approach limits critical reflection on each actor's roles and needs, and it inspires perverse incentives for over-reporting or underreporting, claiming outcomes that are not necessarily yours or quantifying outcomes in ways that do not necessarily provide a meaningful perspective on effectiveness. In the absence of visible change, funding is often considered not to have been well spent, even though, as I have shown in this dissertation, change is not always visible or the only element that should be considered in effectiveness. A shift in focus towards more qualitative evaluation would make evaluation findings more valuable in the sense of shifting from quantity to quality and deepening the understanding of how change actually comes about. This does not necessarily

mean that the measurable component has to be relinquished entirely; there are still certain items that can be measured quantitatively, such as, for example, the number of agenda-setting meetings. Still, though, this information can be better interpreted in terms of meaning when there is information about what the meeting was about, the diverse roles played by the actors involved, how it came about and what interactions led to what kinds of follow-up.

Qualitative information can add meaning in terms of effectiveness, going beyond the mere focus on outcomes and shifting to account for processes and practices as a whole. This can add to learning for evaluators, advocates and policy makers in terms of understanding and doing more justice to advocacy complexity; incorporating critical reflection on underlying assumptions; and developing strategies, plans and donor agreements that are more realistic for the complex and dynamic interventions they support. In the MFS II ILA evaluation, the evaluators constructed learning moments through interactive feedback loops with the evaluated organisations and focused on gathering narratives around how change happened. The findings presented in this dissertation pave the way for a new method of evaluation that does justice to these advocacy dynamics, based on a perspective of advocacy as a socially constructed strategic practice, with advocates acting with strategic agency.

6.4.2. A shift in the international development world

The findings presented in this dissertation represent a call for the development world to step out of the comfort zone of tools, methods and approaches that seek control, prediction, measurability and visibility, and to really focus instead on the meaning of practices and processes including interactions, strategy making, decision making, organisational structures, roles, values, interests and considerations. This necessarily means questioning and reflecting on one's own role—whether you are an advocate, a manager, an evaluator or another actor involved in these processes. It also requires questioning the implicit assumptions, values and norms that shape the current state of affairs, such as the politics of results, which affects the way outcomes are interpreted, measured, presented and given meaning.

This is part of the broader tension between theory and practice (Mosse, 2005; Ramalingam et al., 2008). Theories seek to translate the reality of practices into models that are based on key assumptions about effectiveness, results, how change happens and the roles of diverse actors in achieving change. These theoretical models have become the standard used by practitioners to communicate with the public and donors about the results they pursue and achieve. Again, these models are built on assumptions, which are theoretical understandings of how things work (i.e.

how certain activities or practices should lead to certain changes). Although, in theory, evaluation is objective and rigorous, and results are predictable and controllable, in practice, the reality is much messier. As described in more detail above, twinning the ToC approach with practices of change can pave the way for the integration and interconnection of the different narratives to achieve better interaction between theory (models, policies) and practices (complex, dynamic, strategic).

6.4.3. Steps forward: A future research and practice agenda

Meaningful assumptions

Future research would benefit from a deeper understanding of the role and meaning of the underlying assumptions in evaluation practices, in advocacy and in international development more broadly. The field would benefit from further research that complements and builds on the findings presented in this dissertation, such as activities to enhance the empirical data on advocacy practices and advocacy evaluation. More empirical data would facilitate the development of evaluation methods suitable for assessing such dynamic processes. Additionally, advocacy evaluation methods developed by building on existing knowledge and methods used for monitoring, planning, strategizing and outcome identification would be a welcome development for the field. In this pursuit, two key questions should be taken into account: How can rigour be ensured in assessments building on narratives? What specific approaches work well for capturing advocacy (such as most significant change or contribution analysis), and how should these approaches be adjusted to better suit the diverse purposes of evaluation and the diverse realities of advocacy? One of the issues encountered in the MFS II ILA evaluation involved the relevance of outcomes with regard to the ToC of the intervention, the relevance to the intended beneficiaries, and the relevance in the context of the programme's main goals. Advocacy practices and evaluation will benefit from further study and translation of how relevance can be identified and assessed.

To reduce the tensions between the reality of theory and the reality of practice and to gain a deeper understanding of meaning-making, it is important for advocates, evaluators and donors to question the underlying assumptions of programmes, strategies and practices. What is the role of underlying assumptions, both in advocacy and in its evaluation? How do underlying

assumptions shape interactions, relations and actions? What roles do assumptions play in what advocates achieve, how they strategize and how they interpret their achievements?

Donors and policy makers funding advocacy and seeking to know whether and how the interventions they fund have made a difference should critically reflect on their own assumptions and their own roles in demanding certain results. Based on the findings presented in this dissertation, I believe donors as well as evaluators need to take into account the broader funding context and the existing dependencies in the development world. This includes the critical reflection on and questioning of the assumptions motivating their funding of certain programmes, how they understand change to happen and how they perceive their role in the process. For the evaluations they pursue and plan, it is important to be very clear about the purpose of the evaluation and to translate this purpose into the right evaluation questions. There are limitations to the purposes and the questions asked that should be taken into account.

Building on and connecting to existing approaches

Another issue that came up during my research that is important for practitioners is the fact that, in international development methods, approaches and tools are increasingly being developed to plan, strategize, monitor and evaluate effectiveness. However, these methods, approaches and tools do not always build on the knowledge and methods that already exist. For future research and practice agendas, it is important to include the development of more suitable methods that build on the knowledge that is already available. Currently, newly developed methods too often become hot (e.g. the ToC approach, outcome mapping and outcome harvesting) and are widely implemented by NGOs or demanded by donors in funding proposals. However, although these methods are valuable, they do not provide ideas or tools to ensure that they can be used in cooperation and strengthen each other. As a result, these methods are often used separately, even though connecting their strengths might help in gaining a better understanding of the complex and dynamic processes involved.

Stepping away from outcomes as standard for effectiveness and its measurement

An overarching question that has come up during the entire PhD process concerns outcomes as the standard for measuring effectiveness. During the evaluation and the analysis of the data gathered during the evaluation, this question was on my mind. Should we still use outcomes as the standard for effectiveness and its measurement, or should we start looking at other aspects

that might do more justice to the complexities of the messy world where we live and work? Based on the findings presented in this dissertation, I argue that it is important to let go of the frame that has shaped our idea of effectiveness as being performance- and goal-driven and outcome-focused. This way of thinking has established the convincing perspective that effectiveness is measurable and should be evaluated objectively, and that evaluation that is not objective is not right. Based on my findings, I think we should work towards a new standard of effectiveness, moving away from a narrow focus on outcomes and towards a more inclusive understanding of how effectiveness is given meaning through interactions among actors who operate in systems and structures that they interact with strategically and emotionally. What, then, should an evaluation method include, and what is needed to make a more value and meaningful assessment of effectiveness? These are questions that should be considered by researchers, practitioners, evaluators and donors.

Summary

Negotiating Effectiveness: The politics of Results in advocacy for development and its evaluation

Bodille Arensman

This thesis analyses how advocacy effectiveness is given meaning in the practices of advocacy for development, and how this is affected by the politics of results in its evaluation. Hence, aiming to answer the following research question: *How does the practice of advocacy for development give meaning to effectiveness, and how do the politics of results play out in the evaluation of advocacy?* With advocacy gaining ever more ground in the development world, the issue of (measuring) effectiveness becomes increasingly relevant. Especially when realising that the pursuit and achievement of effectiveness in terms of structural change and development are political processes that require strategies such as advocacy (Hudson, 2002; Jolly et al., 2004; Jones, 2011; Keck and Sikkink, 1998; Kremer et al., 2009; Teles and Schmitt, 2011). There is an assumption among NGOs and donors that advocacy will result in structural changes, thus increasing development effectiveness. Actually proving advocacy for development effectiveness is however, difficult because structural changes in international development often take a very long time to develop, and involve many diverse stakeholders operating at multiple levels and across borders. Besides, advocacy outcomes are often intangible. It is even more complicated to establish the effectiveness of efforts to reach these outcomes.

With advocacy gaining ever more ground in the development world, these issues are becoming increasingly relevant. Advocacy presents challenges for the evaluation of effectiveness, especially because of advocacy's dynamic character and because of the substantive meaning of effectiveness in the contexts where advocates and advocacy evaluators operate (i.e. *politics of results*). In pursuing structural changes, advocates need to navigate, manoeuvre and strategize in changing circumstances. The results of these actions, however, are not necessarily obvious or even visible. How effectiveness is given meaning in the context of such processes is thus challenging for evaluators. Doing this research I found that advocacy evaluation is very much socially and politically constructed. Evaluators and advocates need to understand and act on the fact that knowledge about advocacy effectiveness and practices is

often tacit. Much of what happens is strategically acted upon by advocates, who move in changing environments; negotiate and manoeuvre through interests, needs and demands; and interact with other stakeholders. Consequently, in the field of advocacy, there is no agreed view on 'what works when'.

Meanwhile, effectiveness is pursued and understood as the processes and practices through which an intervention reaches its goals. This influenced the type of models developed to assess effectiveness which focus on prediction and control. This pressures organisations to predefine and show measurable results preferably within funding cycles, leaving its mark on procedures and protocols. This results in that effectiveness often boils down to the question of whether the intervention or organisation has succeeded in achieving legitimacy and credibility in the eyes of other stakeholders. Hence, it is not about asking *whether* an intervention reached its goals, but rather *how* these goals were reached by an intervention. The determination of whether a course of action was effective is not self-evident, and this dissertation understands the meaning of 'effectiveness' to be determined through a process of human interaction that is intrinsically political. This process of defining outcomes and effectiveness is profoundly political. I refer to this political process as the *politics of results*—in which negotiations take place by which outcomes and effectiveness are given meaning.

This doctoral dissertation studied advocacy for development and its evaluation as part of one of the largest transnational advocacy evaluation programmes worldwide. The evaluation measured two moments in time (2012 and 2014), creating a multiple-year programme that enabled the evaluating team—with permission from the donor and stakeholders—to weave a PhD project into the process. Having a dual role of both an evaluator and a PhD researcher, provided for a unique opportunity to observe, follow and study these processes. For this research four cases were studied: three advocacy programs and their evaluation, and the evaluation itself. During the evaluation, a lot of politics were at play in the evaluation, because politics are an inherent part of human interactions—including interactions about conceptualising effectiveness; interactions focused on identifying, assessing or presenting outcomes; and strategic interactions and negotiations that take place as part of evaluation. Indeed, especially around results, there were many interest-driven processes regarding what information was accessed and how the results were identified and presented. Along the way, I learned that the 'honest story' about effectiveness is much more layered and multidimensional than I had thought.

Chapter one examines the puzzle of advocacy for development and its evaluation, setting out the theoretical discussions conceptualizing the politics of results and the traditions of

evaluation including effectiveness. Although the literature provides important insights, it does not touch upon how advocacy effectiveness is given meaning or how this is affected by the politics involved around results, regarding the accountability, credibility and legitimacy of the organisation. There is a gap in the literature pertaining to the specific challenges of the effectiveness of advocacy for development and its measurement.

Chapter two explores how ownership and effectiveness are simultaneously pursued by a transnational advocacy network, examining the tension this brings forward. Ownership and effectiveness are both global policy demands, and both are often included in donor–recipient agreements as part of the funding relations. However, the two concepts continually challenge and define one another. Effectiveness is demanded in measurable, presentable terms, whereas ownership does not readily lend itself to measurement. Moreover, ownership demands a different understanding of effectiveness—as an ambiguous, multidimensional concept that acquires meaning through interactions between social and political considerations. The politics of results in international development frames ownership as being in direct relation with effectiveness. This assumption is questioned, as the findings demonstrate that the relation between the two concepts is ambiguous. This chapter discusses what this tension means for how advocacy effectiveness is shaped through interactions between these two main objectives pursued by the examined network. The interplay between ownership and effectiveness is analysed on three dimensions: collective identity, the accountability process and the advocacy message. The question is not about being more or less effective, but about the processes that shape the meaning of effectiveness in particular institutional constellations.

Chapter three analyses how effectiveness is negotiated in advocacy evaluation. The politics of results affect evaluation theory and practice by putting outcomes at the centre of effectiveness and by assuming these can be objectively identified, measured and presented. The literature includes only a limited discussion of how this influences what is evaluated, how it is evaluated and why is evaluated. Chapter 3 argues that a shift or reconsideration is necessary in the current positivist trends (i.e. performance measurement, results-based management), especially because advocacy effectiveness is not measurable in the rigorous manner that these kinds of methods pursue (i.e. tangible, visible, short-term outcomes). Rather, advocacy evaluation is necessarily subjective. This subjectivity provides for interesting opportunities to probe, reflect on and develop insight into the value of advocacy work. This chapter offers a way to make room and generate understanding of the diverse realities in evaluation. This includes the vested interests involved for evaluators seeking to make a critical assessment and for those being evaluated, who will not always find these assessments to be exactly what they want to hear.

Chapter four questions whether the Theory of Change approach lives up to this expectation of suitability for complex and dynamic interventions such as advocacy. Theory of Change is *hot* in evaluation practice, as well as in the development world, where it is being used and implemented by donors, policy makers and NGOs. This approach is considered especially suitable for complex and dynamic processes. For advocacy evaluation, Theory of Change as evaluation approach remains too stuck in theory, while the recursive, unpredictable and nonlinear nature of advocacy demands more space to account for practices of change. Theory of Change therefore, does not live up to its potential to guide complex processes such as advocacy. It does not provide the necessary tools for evaluators to understand the interaction between theory and practice. Understanding dynamic processes such as advocacy demands 'zooming in' to examine the theory on which strategies and assumptions are based in relation to practices—how advocates act, adapt, manoeuvre and strategize in their everyday realities. Existing work has not necessarily provided a tangible, overarching understanding of what this means for advocacy and its evaluation. This chapter demonstrates that change is initiated from strategies as practices, as it is recursive leading to emergent rather than predicted outcomes, and is shaped through human interaction. The chapter offers an analysis of twinning Theory of Change with Practices of Change, bridging a connection between the theory of evaluation methods and strategy as practice—something actors *do* rather than something organisations *have* (i.e. strategy-as-practice and recursiveness, as seen in organisational studies). This provides space to put advocates at the centre of evaluation, because their practical judgements and manoeuvring of situations and stakeholders shape their interpretations of events that give meaning to effectiveness.

Chapter five further analyses the problematic nature of advocacy outcomes, as intangible, not measurable and untraceable. The identification of outcomes is dependent on the diverse interpretations of the stakeholders involved and advocacy outcomes cannot be considered self-evident. Existing approaches for outcome identification like Outcome mapping and Outcome harvesting do not provide guidelines to deal with the problematic nature of identifying advocacy outcomes. Advocacy outcomes are constructed in multiple layers and are built across borders. This process involves multiple actors, as well as networks that often include advocacy targets. Outcomes are political in terms of how certain outcomes are claimed or reported while others are specifically not claimed. In some cases, outcomes are identified (claimed, not claimed, over-reported or underreported) to fit the purpose of the evaluation (i.e. showing measurable results). In other cases, identifying outcomes and judging effectiveness is more related to the interpreted (diverse realities), constructed (multi-layered) and political (multidimensional, multiple

interests involved) character of outcomes than to some kind of 'objective' measurement. Advocacy outcomes need to be (re-)constructed, questioned and interpreted in an inquisitive way, and evaluators need to pay more attention to the processes and practices through which advocacy effectiveness is given meaning and to the way outcomes are identified.

The final overarching and concluding chapter brings together the discussions and conclusions from the previous chapters. This chapter distils four overarching conclusions and sets out steps forward for future research and an agenda for practitioners (i.e. evaluators, advocates, policy makers). First, in advocacy evaluation there is formal and informal meaning making around effectiveness. Informally, space was created for discussions and learning, revealing the fundamental dynamics, challenges and contexts in which the considerations, negotiations and interpretations underlying strategic decision making were formed. Formally, the evaluated organisations were keen to present their outcomes (results) as more successful (over-reporting, claiming visibility), especially in writing. Second, in these processes the politics of results played an important role and were multidimensional in pursuing, negotiating, understanding and identifying effectiveness. This appears in the form of interests that manifest themselves in struggles between the needs (legitimacy, credibility) and demands (accountability) of both the evaluators and the evaluated. Third, evaluating advocacy therefore demands a different approach. Current evaluation methods do not do enough justice to or support and accommodate the dynamic processes of advocacy. Advocacy evaluation demands an approach that focuses on the strategic practices and practitioners acting strategically. Finally, it is argued that a shift is necessary in international development thinking about effectiveness. A shift away from the comfort zone of tools, methods and approaches that seek control, prediction, measurability and visibility. A shift towards really focusing on the meaning of practices and processes including interactions, strategy making, decision making, organisational structures, roles, values, interests and considerations.

References

African Child Policy Forum (2013). The African report on child wellbeing 2013: Towards greater accountability to Africa's children. Addis Ababa, Ethiopia: ACPF.

Alexander J., Brudney J.L. and Yang K. (2010). Introduction to the symposium: accountability and performance measurement: the evolving role of nonprofits in the hollow state. Nonprofit and Voluntary Sector Quarterly, 39(4): 565-570.

Alliance for Justice (2015). *Philanthropy advocacy playbook: leveraging your dollars*.

Almog-Bar M. and Schmid H. (2013). Advocacy activities of nonprofit human service organisations: a critical review. *Nonprofit and Voluntary Sector Quarterly*, *43*(1): 11-35.

Arendt H. (1998). *The Human Condition. with an introduction by Margaret Canovan*. Chicago and London: University of Chicago Press.

Arensman B., Barrett J., Bodegom van A., Hilhorst D., Klaver D., Rasch E., Richert W., Waegeningh van C., Wessel van M. (2013). *MFS II Joint Evaluation of International Lobbying and Advocacy. Baseline report.* The Hague: NWO.

Arensman B., Barrett J., Bodegom van A., Hilhorst D., Klaver D., Rasch E., Richert W., Waegeningh van C., Wagemakers A., Wessel van M. et al. (2015). *MFS II Joint Evaluation of International Lobbying and Advocacy*. The Hague: NWO.

Arensman B. and van Wessel M. (2017). Negotiating effectiveness in transnational advocacy evaluation. *Evaluation* (first published online 4 October 2017), *24* (1): 51-68.

Arensman B., Wessel M. van, Hilhorst D. (2017a). Does local ownership bring about effectiveness? The case of a transnational advocacy network. *Third World Quarterly*, *38*(6), 1310-1326.

Arensman, B., van Waegeningh, C., and van Wessel, M. (2017b). Twinning "Practices of Change" With "Theory of Change" Room for Emergence in Advocacy Evaluation. *American Journal of Evaluation*, 1 September: 1-16.

Armytage L. (2011). Evaluating aid: an adolescent domain of practice. *Evaluation, 17* (3): 261–276.

Bamberger M., Tarsilla, M. and Hesse-Biber S. (2016). Why so many "rigorous" evaluations fail to identify unintended consequences of development programs: How mixed methods can contribute. *Evaluation and Program Planning*, *55*: 155–162.

Barnes M., Matka E., and Sullivan H. (2003). Evidence, understanding and complexity: Evaluation in non-linear systems. *Evaluation*, *9* (3): 165–284.

Barrett J., Wessel van M. and Hilhorst D.. (2016). *Advocacy for Development: Effectiveness, Monitoring and Evaluation*. Wageningen: Wageningen University.

Beer T. and Reed E. D. (2009). A model for multilevel advocacy evaluation. *Innovation Network Inc, 1* (3): 149–161.

Best J. (2017). The rise of measurement-driven governance: the case of international development. *Global Governance*, *23*: 163-181.

Betterevaluation (2017). *Outcome Harvesting*. Published online: http://www.betterevaluation.org/en/plan/approach/outcome_harvesting (accessed 9 July 2017)

Binderkrantz A. (2005). Interest group strategies: navigating between privileged access and strategies of pressure. *Political Studies*, *53* (4): 694-715.

Booth D. (2012). Aid Effectiveness: Bringing Country Ownership (and Politics) Back in. *Conflict, Security and Development, 12* (5): 537–558.

Bourdieu P. (1990). *The logic of practice*. Stanford University Press.

Brinkerhoff D. W. (2008). The state and international development management: Shifting tides, changing boundaries, and future directions. *Public Administration Review*, *68* (6), 985-1001.

Brown L. D., Ebrahim A., and Batliwala S. (2012). Governing international advocacy NGOs. *World Development*, *40* (6), 1098-1108.

Brown W. A. (2005). Exploring the association between board and organizational performance in nonprofit organizations. *Nonprofit Management and Leadership, 15* (3): 317-339.

Calhoun C. J., and McGowan J. (Eds.). (1997). *Hannah Arendt and the meaning of politics* (Vol. 6). U of Minnesota Press.

Canovan M. (1994). *Hannah Arendt: a reinterpretation of her political thought*. Cambridge University Press.

Carbert A. (2004). Learning from experience: activist reflections on 'insider-outsider' strategies. *AWID (Association of Women's Rights in Development) Spotlight*, (4).

Chambers R. (1995). The Professionals and the Powerless: Whose Reality Counts. *World Summit for Social Development, 4* (1): 14–15.

Chapman J. and Wameyo A. (2001). *Monitoring and evaluating advocacy: a scoping study*. UK: Action Aid.

Coburn C. E., and Penuel W. R. (2016). Research–practice partnerships in education: Outcomes, dynamics, and open questions. *Educational Researcher, 45* (1): 48–54.

Coffman J. (2007). What's different about evaluating advocacy and policy change. *The Evaluation Exchange*, *13* (1): 2-4.

Coffman J. (2009). Overview of current advocacy evaluation practice. Available at: www.evaluationinnovation.org/sites/default/files/Coffman%20Brief%201.pdf (accessed 9 July 2017).

Coffman J. (2011). Revolutionary lessons about interim outcomes. *Advocacy evaluation update, 11* (summer).

Coffman J. and Reed E. (2009). Unique Methods in Advocacy Evaluation. Los Angeles: The California Endowment. *Retrieved From*: http://www.pointk.org/resources/files/Unique_Methods_Brief.pdf

Corner A., and Hahn U. (2010). Message framing, normative advocacy and persuasive success. *Argumentation*, *24* (2): 153-163.

Coryn C. L., Noakes L. A., Westine C. D., and Schröter D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation*, *32* (2), 199–226.

Crozier M. (2007). Recursive governance: Contemporary political communication and public policy. *Political Communication*, *24* (1), 1–18.

Dahl R. (1964). *Power, pluralism and democracy: a modest proposal*. New Haven, CT: Yale University Press.

Denton R. E. Jr. and Woodward G.C. (1998). *Political communication in America.* (3rd edition). Westport CT and London: Praeger.

Devlin-Foltz D., Fagen M.C., Reed E. et al. (2012). Advocacy evaluation: challenges and emerging trends. *Health Promotion Practice, 13* (5): 581–586.

DFID UK Department for International Development. (2011). *DFID's Approach to Value for Money*. London: July. Accessed December 1 2016: https://www.gov.uk/government/uploads/system/uploads/ (accessed 9 July 2017).

Donais T. (2009). Empowerment or Imposition? Dilemmas of Local Ownership in Post-Conflict Peacebuilding Processes. *Peace and Change, 34* (1): 3–26.

Eade D. (Ed.). (2002). *Development and culture* (p. ix). Oxford, United Kingdom: Oxfam.

Earl S., Carden F., and Smutylo T. (2001). *Outcome Mapping: Building Learning and Reflection into Development Programmes*. Ottawa: IDRC.

Easterly W. (2006). *The white man's burden. Why the West's efforts to aid the rest have done so much ill and so little good*. New York: The Penguin Press.

Ebrahim A. (2003a). Making Sense of Accountability: Conceptual Perspectives for Northern and Southern Nonprofits. *Nonprofit Management and Leadership, 14* (2): 191–212.

Ebrahim A. (2003b). Accountability in practice: mechanisms for NGOs. *World Development, 31*(5): 813-829.

Edwards M. and Hulme D. (1995). *Non-governmental organisations: performance and accountability beyond the magic bullet*. Earthscan.

Edwards M. and Hulme D. (1996). Too close for comfort? The impact of official aid on nongovernmental organisations. *World Development*, *24* (6): 961-973.

Emirbayer M., and Johnson, V. (2008). Bourdieu and organizational analysis. *Theory and society*, *37* (1): 1-44.

Es van M. and Guijt I. (2015). Theory of change as best practice or next trick to perform? Hivos' journey with strategic reflection. In R. Eyben, I. Guijt, C. Roche, and C. Schutt (Eds.) *The politics of evidence and results in international development: Playing the game to change the rules?*: 95-114. Rugby, UK: Practical Action Publishing.

Eyben R. (2008). Power, mutual accountability and responsibility in the practice of international aid: A relational approach. *Institute of Development Studies.* Working Paper 305.

Eyben R. (2010). Supporting Inclusive and Democratic Ownership. a 'How to Note' for Donors. *OECD DAC*. Consultation: 29–30 November. Accessed December 1, 2016.

http://cso-effectiveness.org/IMG/pdf/enabling_environment_institute_for_development_studies.pdf

Eyben R. (2012). Theory of Change: breaking out of the results agenda. *Online discussion Big Push Forward*. 11 March: http://bigpushforward.net/archives/1419 (accessed 16 July 2012).

Eyben R. (2015). Uncovering the politics of evidence and results. In R. Eyben, I. Guijt, C. Roche, and C. Schutt (Eds.), *The politics of evidence and results in international development: Playing the game to change the rules?*: 19-38. Rugby, UK: Practical Action Publishing.

Eyben R. and Savage L. (2013). Emerging and Submerging Powers Imagined Geographies in the New development partnership at the Busan fourth high level forum. *The Journal of Development Studies*, *49* (4): 457-469.

Eyben R., Guijt I., Roche C. and Schutt C. (Eds.) (2015). *The politics of evidence and results in international development: Playing the game to change the rules?* Rugby, UK: Practical Action Publishing.

Eyben R., Guijt I., Roche C. et al. (eds) (2015). *The Politics of Evidence and Results in International Development: Playing the Game to Change the Rules?* Rugby: Practical Action Publishing.

Fagen, M. C., Reed, E., Kaye, J. W., and Jack Jr, L. (2009). Advocacy evaluation: What it is and where to find out more about it. *Health promotion practice, 10* (4): 482-484.

Faust J. (2010). Policy Experiments, Democratic Ownership and Development Assistance. *Development Policy Review, 28* (5): 515–534.

Feldman M. S. and Orlikowski W. J. (2011). Theorizing practice and practicing theory. *Organization Science, 22* (5): 1240–1253.

Fowler A. (2000). *Civil Society, NGDOs and Social Development: Changing the Rules of the Game*. Geneva: United Nations Research Institute for Social Development.

Fyall R. and MacGuire M. (2015). Advocating for policy change in non-profit coalitions. *Nonprofit and Voluntary Sector Quarterly, 44* (6): 1274-1291.

Gamson J. (1997). Messages of Exclusion: Gender, Movements, and Symbolic Boundaries. *Gender and Society, 11*: 178–199.

Gardner A. and Brindis C. (2017). *Advocacy and policy change evaluation: theory and practice*. US: Stanford. Stanford University Press.

Giddens A. (1984). *The constitution of society: outline of the theory of structuration*. UK: Cambridge: Policy Press.

Global Partnership for the Prevention of Armed Conflict. (2011). GPPAC Strategic plan 2011-2015. The Hague: GPPAC. January.

Gormley Jr. W. T., and Cymrot H. (2006). The strategic choices of child advocacy groups. *Nonprofit and Voluntary Sector Quarterly*, *35* (1): 102-122.

Grantcraft (2012). Speaking up! Foundations and Advocacy in Europe. *Practical wisdom for funders series*. Belgium: Brussels.

Guba E.G. and Lincoln Y. (1989). *Fourth Generation Evaluation*. Berkeley, CA: Sage Publications.

Guba E.G. and Lincoln Y. (1994). Competing paradigms in qualitative research. In: Denzin N.K. and Lincoln Y.S. (eds). *Handbook of Qualitative Research.* London: Sage: 105–117.

Gulrajani, N. (2011). Transcending the great foreign aid debate: managerialism, radicalism and the search for aid effectiveness. *Third world quarterly*, *32* (2): 199-216.

Guthrie K., Louie J., David T., and Foster C. C. (2005). The challenge of assessing policy and advocacy activities: Strategies for a prospective evaluation approach. *The California Endowment*: 1-56.

Harvard Family Research Project (2007). The evaluation exchange. A periodical on emerging strategies in evaluation. *Harvard Graduate School of Education*. XIII(1) Spring: retrieved from: http://www.hfrp.org/var/hfrp/storage/original/application/6bdf92c3d7e970e7270588109e23b678.pdf (accessed 9 August 2012).

Hatton M. J., and Schroeder K. (2007). Results-based management: friend or foe?. *Development in Practice*, *17* (3), 426-432.

Hayman R., King S., Kontinen T. et al. (2016). *Negotiating Knowledge: Evidence and Experience in Development NGOs*. Rugby: Practical Action Publishing.

Henry L., Mohan G. and Yanacopulos H. (2004). Networks as Transnational Agents of Development. *Third World Quarterly, 25* (5): 839–855.

Herman R. D., and Renz D. O. (1997). Multiple constituencies and the social construction of nonprofit organization effectiveness. *Nonprofit and voluntary sector quarterly, 26* (2): 185-206.

Herman R. D., and Renz D. O. (1999). Theses on nonprofit organizational effectiveness. *Nonprofit and voluntary sector Quarterly*, *28* (2): 107-126.

Hilhorst D. (2003). *The real world of NGOs: Discourses, diversity and development*. London, UK: ZedBooks Ltd.

Hoskyns C. (1999). Gender and Transnational Democracy: The Case of the European Union. In *Gender Politics in Global Governance*, edited by M. K.Meyer and E.Prugl. Lanham: Bowman and Littlefield: 72–87.

Hudson A. (2001). NGOs' transnational advocacy networks: from 'legitimacy' to 'political responsibility? *Global Networks, 1* (4): 331–352.

Hudson A. (2002). Advocacy by UK-based development NGOs. *Nonprofit and Voluntary Sector Quarterly*, *31* (3): 402-418.

Hulme D. (2010). Lessons from the making of the MDGS: human development meets results-based management in an unfair world. *IDS bulletin, 41* (1): 15-25.

Hulscher M. E. J. L., Laurant M. G. H. and Grol R. P. T. M. (2003). Process evaluation on quality improvement interventions. *Quality and Safety in Health Care*, *12* (1): 40-46.

Hunt S. A. and Benford R. D. (2004). The Collective Identity, Solidarity and Commitment. In *The Blackwell Companion to Social Movements*, edited by D. A. Snow, S. A. Soule and H Kriesi. Oxford: Blackwell Publishing: 433–457.

Independent Commission for Aid Impact (2011). *ICAI's Approach to Effectiveness and Value for Money*. Report 1, November, UK.

Innovation Network (2007). *Advocacy Evaluation Update*. 1, June. Washington, D.C.

Innovation Network (2008). *Speaking for themselves: advocates' Perspectives on Evaluation*. Washington D.C.

Innovation Network (2012). *State of Evaluation: Evaluation practice and capacity in the nonprofit sector*. Washington, D.C.

IOB Policy and Operations Department (2009). *Evaluation Policy and Guidelines for Evaluations*. The Hague: Netherlands Ministry of Foreign Affairs.

Jabeen S. (2016). Do we really care about unintended outcomes? An analysis of evaluation theory and practice. *Evaluation and Program Planning*, *55*: 144–154.

James, C. (2011). Theory of change review. *Comic Relief*.

Jarzabkowski P. (2004). Strategy-as-practice: Recursiveness and adaptation, and practices-in-use. *Organization Studies*, *25* (4): 529–560.

Jarzabkowski P. (2005). *Strategy as practice: An activity-based approach.* London, UK: Sage.

Jasper J., and Polletta F. (2001). Collective Identity and Social Movements. *Annual Review of Sociology, 27* (1): 283–305.

Jolly R., Emmerij L., Ghai D. and Lapeyre F. (2004) UN Contributions to development thinking and practice. *United Nations Intellectual History Project Series*. Bloomington: Indiana University Press.

Jones H. (2011). *A guide to monitoring and evaluating policy influence* (ODI Background note). London, UK: Overseas Development Institute.

Jones H., and Hearn S. (2009). *Outcome mapping: A realistic alternative for planning, monitoring and evaluation*. Overseas Development Institute.

Jordan L., and Tuijl van P. (2000). "Political Responsibility in Transnational NGO Advocacy." *World Development, 28* (12): 2051–2065.

Katz H., and Anheier H. (2005). Global Connectedness: The Structure of Transnational NGO Networks. *Global Civil Society, 6*: 240–265.

Keck M. and Sikkink K. (1998). *Activists Beyond Borders*: *Advocacy Networks in International Politics.* Ithaca, NY and London: Cornell University Press.

Keck M. E., and Sikkink K. (1999). Transnational advocacy networks in international and regional politics. *International social science journal*, *51* (159): 89-101.

Kerkvliet B. J. (2009). Everyday politics in peasant societies (and ours). *The journal of peasant studies, 36* (1): 227-243.

Klugman B. (2011). Effective social justice advocacy: A theory-of-change framework for assessing progress. *Reproductive Health Matters, 19* (38): 146–162.

Knaap van der P. (2004). Theory-based evaluation and learning: possibilities and challenges. *Evaluation, 10* (1): 16–34.

Korten D. C. (1987). "Third Generation NGO Strategies: A Key to People-Centered Development." *World Development, 15*: 145–159.

Kremer M., Liehout van P. and Went R. (eds.) (2009) *Doing Good or Doing Better. Development policies in a globalizing world*. WRR: Scientific Council for Research Policy. Amsterdam: Amsterdam University Press.

Lam T.C.M. (2002). Theory-based evaluation and objective-based evaluation: an integration of the two approaches. In: *European Evaluation Society conference*. Seville, Spain, 12 October.

Langley A. (2007). Process thinking in strategic organization. *Strategic Organization*, *5* (3): 271–282.

Lemon, A., and Pinet, M. (2017). Measuring unintended effects in peacebuilding: What the field of international cooperation can learn from innovative approaches shaped by complex contexts. *Evaluation and Program Planning*.

Liket K.C., Rey-Garcia M., and Maas K.E.H. (2014). Why aren't evaluations working and what to do about it: a framework for negotiating meaningful evaluation in nonprofits. *American Journal of Evaluation, 35* (2): 171–188.

Lo A., and Mueller M. (2010). WARNING: Physics envy may be hazardous to your wealth!.

Long N. (1989). *Encounters at the interface. A perspective on social discontinuities in social life*. The Netherlands: Wageningen Agricultural University.

Long N. (2003). *Development sociology: Actor perspectives*. London, UK: Routledge.

Louie J., and Guthrie K. (2007). Strategies for assessing policy change efforts: A prospective approach. *The Evaluation Exchange*, *13* (1): 5–7.

Lowery, D. (2013). Lobbying influence: Meaning, measurement and missing. *Interest Groups and Advocacy*, *2* (1): 1-26.

Lynn J. (2014). *Assessing and evaluating change in advocacy fields*. Washington, D.C.: Center for Evaluation Innovation.

Mansoor K. (2003). Realist evaluation for practice. *British Journal of Social Work, 33*: 803–818.

Marchal B., Belle van S., and Olmen van J. (2012). Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. *Evaluation, 18* (2): 192–212.

Marcus G. E. (1995).Ethnography in/of the World System: The Emergence of Multi-Sited Ethnography. *Annual Review of Anthropology*: 95–117.

Markiewicz A. (2005). 'A balancing act': resolving multiple stakeholder interests in program evaluation. *Evaluation Journal of Australasia, 5* (1and2): 13–21.

Marston S. (2000). The Social Construction of Scale. *Progress in Human Geography, 24* (2): 219–242.

Mason P., and Barnes M. (2007). Constructing theories of change: Methods and sources. *Evaluation, 13* (2): 151–170.

Mayne J. (2008). Contribution analysis: An approach to exploring cause and effect. *ILAC Brief*, (16) Mai.

Mayne J. (2012). Contribution analysis: Coming of age?. *Evaluation*, *18* (3): 270-280.

McDonald D. (2008). Revisiting a theory of negotiation: the utility of Markievicz (2005) proposed six principles. *Evaluation and Program Planning, 31*: 259–265.

Morariu J., and Brennan K. (2009). Effective advocacy evaluation: The role of funders. *The Foundation Review*, *1* (3): 100–108.

Mosley J. E. (2011). Institutionalization, privatization, and political opportunity: What tactical choices reveal about the policy advocacy of human service nonprofits. *Nonprofit and Voluntary Sector Quarterly*, *40* (3): 435-457.

Mosse D. and Lewis D. (2005). *The aid effect: giving and governing in international development*. London, UK: Pluto Press.

Mouffe C. (2005). *The return of the political* (Vol. 8). New York: Verso.

Mowles C. (2013). Evaluation, complexity, uncertainty: Theories of change and some alternatives. In T. Wallace, F. Porter, and M. Ralph-Bowman (Eds.), *Aid, NGOs and the realities of women's lives: A perfect storm*. Rugby, UK: Practical Action Publishing: 47–60.

Natsios A. (2010). The clash of the counter-bureaucracy and development. *Centre for Global Development*. US.

Netherlands Ministry of Foreign Affairs (2012). Terms of Reference. MFS II Evaluation on International Lobbying and Advocacy. The Hague.

Netherlands Ministry of Foreign Affairs (2015). *Policy framework: Dialogue and Dissent: strategic partnerships for lobbying and advocacy.* The Hague: The Netherlands: AVT14/BZ111606A, 13-05-2014:https://www.government.nl/documents/regulations/2014/05/13/policy-framework-dialogue-and-dissent

Nownes, A. J., and Newmark, A. J. (2016). The information portfolios of interest groups: An exploratory analysis. *Interest Groups and Advocacy*, *5* (1): 57-81.

Ongevalle van J. and Peels R. (2012). The outcome mapping usefulness barometer. How useful is outcome mapping to help deal with complex change? *Outcome Mapping Learning Community*. KU Leuven and HIVA. January.

Organization for Economic Co-operation and Development. (1991, 1995). *Criteria for Evaluating Development Assistance*. Paris: OECD.

Organization for Economic Co-operation and Development. (2005). *The Paris Declaration on Aid Effectiveness*. Paris: OECD.

Organization for Economic Co-operation and Development. (2008). *The Accra Agenda for Action*. Paris: OECD.

Organization for Economic Co-operation and Development. (2011). *Busan Partnership for Effective Development Co-Operation. Fourth High Level Forum on Aid Effectiveness*. Paris: OECD.

Patton M. Q. (1997). *Utilization-focused evaluation*. Thousand Oaks, CA: Sage.

Patton M.Q. (2011). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York, NY: Guilford Press.

Pawson R. and Tilly N. (1997). *Realistic Evaluation*. London: Sage.

Porter S. (2015). The uncritical realism of realist evaluation. *Evaluation, 21* (1): 65-82.

Prinsen G., and Nijhof S. (2015). Between logframes and theory of change: reviewing debates and a practical experience. *Development in Practice*, *25* (2), 234-246.

Prinsen G., and Nijhof S. (2015). Between logframes and theory of change: Reviewing debates and a practical experience. *Development in Practice*, *25* (2): 234–246.

Provan K. G. and Lemaire R. H. (2012). Core concepts and key ideas for understanding public sector organisational networks: using research to inform scholarship and practice. *Public Administration Review*. September and October.

Provan K. G., and Kenis P. (2008). Modes of Network Governance: Structures, Management and Effectiveness." *Journal of Public Administration and Research, 18* (2): 229–252.

Provan K. G., Fish A. and Sydow J. (2007). Interorganizational Networks at the Network Level: A Review of the Empirical Literature on Whole Networks." *Journal of Management, 33* (3): 479–516.

Putnam L.L. (2004). Transformations and critical moments in negotiations. *Negotiation Journal, 20* (2): 275–295.

Ramalingam B., Jones H., Reba T., and Young J. (2008). *Exploring the science of complexity: Ideas and implications for development and humanitarian efforts* (Vol. 285). UK: London. Overseas Development Institute.

Raynor J., York P. and Sim S.C. (2009). What makes an effective advocacy organization? A framework for determining advocacy capacity. *TCC Group. The California Endowment.*

Reich H. (2006). 'Local Ownership' in Conflict Transformation Projects. Partnership, Participation or Patronage?. *Berghof Occasional Paper, 27* (September): 1–41.

Reisman J., Gienapp A., and Stachowiak S. (2007). *A guide to measuring advocacy and policy*. Baltimore: Organisational Research Services.

Richmond O. P. (2012). Beyond Local Ownership in the Architecture of International Peacebuilding. *Formerly Global Review of Ethnopolitics, 11* (4): 354–375.

Riddell R. (2007). *Does foreign aid really work?* UK: Oxford University Press.

Riddell R.C. (2014). Does foreign aid really work? (keynote address) at the: *Australasian aid and international development workshop*. Canberra. (13 February).

Ringsing B. and Leeuwis C. (2008). Learning about advocacy: a case-study of challenges, everyday practices and tensions. *Evaluation, 14* (4): 413–436.

Roche C. and Kelly L. (2012a). *The evaluation of politics and the politics of evaluation*. Background paper 11, Developmental Leadership Program: Policy and Practice for Developmental Leaders, Elites and Coalitions, (August).

Roche C. and Kelly L. (2012b). *Monitoring and evaluation when politics matter*. Background paper 12, Developmental Leadership Program: Policy and Practice for Developmental Leaders, Elites and Coalitions, (October).

Rogers P. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation, 14*(1): 29–48.

Sayer A. (2000). *Realism and Social Science*. London: Sage.

Sayer A. (2011). *Why Things Matter to People: Social Science, Values and Ethical Life*. Cambridge: Cambridge University Press.

Saz-Carranza A. (2012). *Uniting Diverse Organizations. Managing Goal-Oriented Advocacy Networks*. New York: Routledge.

Saz-Carranza A. and Ospina S. M. (2011). The Behavioural Dimension of Governing Interorganizational Goal- Directed Networks. Managing the Unity-Diversity Tension. *Journal of Public Administration Research and Theory, 21*: 327–365.

Schaap A. (2007). Political theory and the agony of politics. *Political Studies Review*, *5* (1): 56-74.

Sharkey S. and Sharples A. (2008). From the beginning: negotiation in community evaluation. *Evaluation, 14* (3): 363–380.

Shawki N. (2010. Political opportunity structures and the outcomes of transnational campaigns: a comparison of two transnational networks. *Peace and Change, 35* (3) July: 381-411.

Shawki N. (2011). Organizational Structure and Strength and Transnational Campaign Outcomes: A Comparison between Two Transnational Advocacy Networks. *Global Networks, 11* (1): 97–117.

SIDA (2016). Policy framework for Swedish development cooperation and humanitarian assistance 2016/17:60. *Swedish Ministry of Foreign Affairs*. Online through the following link: http://www.government.se/49a184/contentassets/43972c7f81c34d51a82e6a7502860895/skr-60-engelsk-version_web.pdf (accessed 11 September 2017).

Skene-Pratt E. (2013). Nonprofit advocacy: a Michigan primer. *Lansing MI: A joint publication of the Michigan Nonprofit Association and Council of Michigan Foundations.*

Sloot H. and Gaanderse M. (2010). *Guidelines on lobby and advocacy*. ICCO. Den Haag: Drukkerij Roos en Roos.

Smutylo, T. (2005). Outcome mapping: a method for tracking behavioural changes in development programs. ILAC Brief 7 (August).

Stachowiak S. (2013). Pathways for change: 10 theories to inform advocacy and policy change efforts. *Center for Evaluation Innovation*. Seattle, WA: ORS Impact.

Starling, S. (2010). Monitoring and evaluating advocacy: lessons from Oxfam GB's Climate Change campaign. *Development in Practice*, *20* (2): 277-286.

Stein D. and Valters C. (2012). *Understanding theory of change in international development: A review of existing knowledge* (JSRP Paper 1). London, UK: JSRP and The Asia Foundation.

Taylor D. and Balloch S. (2005). The politics of evaluation: an overview. In: Taylor D. and Balloch S. (eds) *The Politics of Evaluation: Participation and Policy Implementation*. Bristol: Policy Press: 1–17.

Teles S. and Schmitt M. (2011). The elusive craft of evaluating advocacy. *Stanford Social Innovation Review,* (Summer): 39–43.

Thomas H., and Wilson A. D. (2011). 'Physics envy', cognitive legitimacy or practical relevance: dilemmas in the evolution of management research in the UK. *British Journal of Management*, *22* (3): 443-456.

Tsui and Lucas (2013). Methodologies for measuring influence. *GSDRC Applied Knowledge Services prepared for DFID*. UK: London.

Tsui J., Hearn S. and Young J. (2014). Monitoring and evaluation of policy influence and advocacy. Working paper 395. *Overseas Development Institute*. London, UK.

Unicef (2010). Advocacy toolkit. A guide to influencing decisions that improve children's lives. New York.

United Nations General Assembly. (2015). *Transforming Our World: The 2030 Agenda for Sustainable Development*. A/RES/70/1 (21 October).

Vaara A., and Whittington R. (2012). Strategy-as-practice: Taking social practices seriously. *The Academy of Management Annals*, *6* (1), 285–336.

Vähämäki, J., Schmidt, M., Molander, J. (2011). Review of Results Based Management in Development Cooperation. Riksbankens Jubileumsfond, Sweden.

Valters C. (2014). *Theories of change in international development: Communication, learning or accountability?* JSRP Paper 17. London, UK: JSRP and The Asia Foundation.

Vogel I. (2012). *Review of the use of 'theory of change' in international development*. UK Department of International Development.

Wallace T. (Ed.) (2006). *The aid chain: Coercion and commitment in development NGOs*. Rugby, UK: Practical Action Publishing.

Weiss C. H. (1997). Theory-based evaluation: past, present and future. *New Directions for Program Evaluation, 76*: 41-55.

Weiss C. H. (2000). Theory-based evaluation: Theories of change for poverty reduction programs. In O. N. Feinstein and R. Picciotto (Eds.) *Evaluation and poverty reduction*. Washington DC: World Bank Publications: 103–114.

Wessel van M., Buuren van R., and Woerkum van C. (2011). Changing planning by changing practice: How water managers innovate through action. *International Public Management Journal*, *14* (3): 262–283.

Whelan J. (2008). *Advocacy evaluation: review and opportunities*. Brisbane, AU: The Change Agency.

Wilson-Grau R. and Britt H. (2012). Outcome Harvesting. *Cairo: Ford Foundation*.

Wilson-Grau, R. (2015). *Outcome Harvesting*. Better Evaluation. Retrieved from http://betterevaluation.org/plan/approach/outcome_harvesting

Witko, C., and Newmark, A. J. (2005). Business mobilization and public policy in the US states. *Social Science Quarterly*, *86* (2): 356-367.

Wong W. H. (2012). *Internal Affairs. How the Structure of NGOs Transforms Human Rights*. Ithaca and London: Cornell University Press.

World Bank. (2013). World Bank, Global Fund, Partner to Expand Results-Based Financing for Maternal and Child Health. Press release: 11 December.

Yanow D. (1993). The communication of policy meanings: Implementation as interpretation and text. *Policy sciences, 26* (1): 41-61.

Yogita M., Johannes L. and Kumar G. (2010). *Output-Based Aid: Lessons Learned and Best Practices*. World Bank.

Zandniapour L. and Brennan K. (2010). Advocacy evaluation and rigor. *Advocacy Evaluation Update, 10*: 1–3.

Annex 1

Executive summary to the evaluation findings of the Joint MFS II Evaluation on International Lobbying and Advocacy

Summary from MFS II ILA Evaluation report findings 2015 by Arensman B., Barrett J., Bodegom van A., Hilhorst D., Klaver D., Rasch E., Richert W., Waegeningh van C., Wagemakers A., Wessel van M. et al. (2015). *MFS II Joint Evaluation of International Lobbying and Advocacy*. The Hague: NWO.

Background to the evaluation

The Co-Financing System (Medefinancieringsstelsel, or 'MFS') is the 2011–2015 grant framework for Co-Financing Agencies (CFA). A total of 20 Alliances of Dutch CFAs were awarded €1.9 billion through the MFS II grants framework by the Dutch Ministry of Foreign Affairs (NL MoFA). The MFS II framework stipulates that each consortium carry out independent external evaluations to be able to make valid, evaluative statements about the effective use of the available funding. To meet these evaluation requirements, a joint evaluation programme was developed and approved by the NL MoFA. The overall purpose for evaluating MFS II-funded development interventions is to account for results and to contribute to the improvement of future development interventions.

The MFS II has been evaluated through country studies. In addition, this evaluation of International Lobby and Advocacy (ILA) was commissioned as a thematic evaluation across the MFS II. This evaluation concerned the lobby and advocacy programmes of 8 Alliances. The specific aims of this ILA programme evaluation are 1) to assess the effectiveness, efficiency and relevance of ILA programmes funded by MFS II; 2) to develop and apply innovative methodologies for the evaluation of ILA programmes and 3) to provide justified recommendations that enable Dutch CFAs and/or their Southern partners to draw lessons for future development interventions. The evaluation covers the period 2011–2014.

The five main research questions have been formulated as follows:

What are the changes achieved in the three priority result areas through international lobbying and advocacy on the thematic clusters 'sustainable livelihoods and economic justice', 'sexual and reproductive health and rights' and 'protection, human security and conflict prevention' during the 2011–2014 period? Do the international lobbying and advocacy efforts

of the MFS II Alliances and their partners contribute to the identified changes (i.e. measuring effectiveness)? What is the relevance of these changes? Were the efforts of the MFS II Alliances efficient? What factors explain the findings drawn from the questions above?

The Call for Proposals for this evaluation (issued by NWO-WOTRO), distinguished three priority result areas for this evaluation: 1) agenda setting, 2) policy influencing and 3) changing practice.

Scope of the evaluation

The scope of the evaluation is determined by the thematic clusters and the priority result areas defined by NWO-WOTRO in the call for proposals. NWO-WOTRO distinguishes three thematic clusters for this evaluation: (i) sustainable livelihoods and economic justice, (ii) sexual and reproductive health and rights and (iii) protection, human security and conflict prevention. Under the responsibility of Partos and WOTRO, Partos' evaluation manager and two consultants developed and carried out the selection of programmes to be evaluated, in consultation with the Alliances. The selection process primarily focused on representativeness in terms of thematic focus. The three clusters were established so that each of the eight Alliances with an ILA component could be represented in the evaluation in a theme that was prominent in their work. In consultation with the Alliances, a programme was selected for evaluation for each Alliance.

Thematic cluster	*Alliances*
Sustainable livelihoods and economic justice	Ecosystems Alliance
	Fair, Green and Global Alliance
	IMPACT Alliance
	Hivos Alliance, People Unlimited 4.1
Sexual and reproductive health and rights	SRHR Alliance
Protection, human security, and conflict prevention	Communities of Change
	Freedom from Fear
	Together4change

The scope of the selected programmes varied substantially across Alliances. In some cases, as is elaborated further in the cluster and Alliance chapters, evaluation sub-teams have had to place additional limits on the boundaries of the selected programmes for reasons of feasibility.

Conceptualising 'lobby and advocacy'

Lobby and advocacy in international development interventions can be defined as a 'wide range of activities conducted to influence decision makers at different levels' towards the overall aim of development interventions to combat the structural causes of poverty and injustice. The concept of advocacy, however, goes beyond mere policy influence and aims for sustainable changes in public and political arenas, including awareness raising, litigation (legal actions) and public education, as well as building networks, relationships and capacity. Methods used to influence decision makers in this regard are 1) persuasion and cooperation (lobby) and 2) pressure (e.g. blaming and shaming) and confrontation (used in other advocacy strategies). Hence, lobby can be understood as one of the strategies for policy influencing, thus advocacy. Lobby is the influencing of policy makers by building relations, creating awareness and finding connections to build enthusiasm among policy makers for the chosen aim. Advocacy also influences decision makers in and through other arenas or channels, such as civil society, the broader public, the private sector and politics. Awareness raising and information sharing can be seen as key strategies to reach decision makers in these arenas.

The importance of networked advocacy was a theme that crosscut the evaluation questions as well as the three thematic clusters in this evaluation. Almost all of the Alliances in the evaluation accomplish their work through networks, but there is variation in the intensity and manner of cooperation. We focused special attention on identifying the role of networking and networked advocacy in the Alliances' ToCs, overall missions and strategic approaches.

Advocacy interventions are oriented towards governmental, political and private sector actors like multinational corporations. Advocacy often seeks to take a systemic perspective and targets the various related levels of influence: supranational, international, interregional, national and local. Advocacy directed at system level changes has been increasingly embraced by non-profit organisations, responding to the public and political calls for results that are more tangible and for accountability on public-funded development interventions.

Advocacy is pre-eminently a field where rapid changes in public, private and political arenas happen, with unexpected interactions, feedback loops and emergent systems generating both opportunities and threats. In this dynamic environment, the advocate moves between multiple layers of relationships and objectives, adapting and finding pathways that are nonlinear in nature, often using an extended time span to ultimately achieve results.

The three priority result areas of agenda setting, policy influencing and changing practice rarely take place in a linear fashion. Although they are often not consecutive, the priority result

areas are seen as core to advocacy. Agenda setting is linked to strategic awareness raising in the public and private sector and in political arenas. Policy influencing focuses on creating a public constituency and changing public and political debate, leading to demonstrable change in policy by the lobby targets. The ultimate aim is to change practice, where the change in policy is realised in reaching development objectives.

Approach and methods

This evaluation set out to do justice to the complexities of international lobbying and advocacy. Our methodological approach meant that we kept an open eye for the flexibility required for lobbying and advocacy, the multiplicity of relations involved, the long-term orientation needed when it comes to seeking change and the highly political nature of the work, with multiple forces often acting against one's objectives.

To answer the first evaluation question ('what are changes achieved'), we have focused our research on changes to which the Alliances claim to have contributed, allowing for consistency with evaluation questions 2–5. We sought to identify all such changes and to explain how the changes related to the three priority result areas, to the Alliances' own ToCs, to the objectives of the programmes under evaluation and to developments in the broader context and thematic and policy focus area of the Alliances. The identification of relevant outcomes was accomplished in cooperation with the Alliances. We have included all outcomes in the period 1 January 2011–1 October 2014.

To answer the second evaluation question ('do the ILA efforts contribute to the identified changes') the evaluation team sought to determine whether a credible (plausible) case can be made that the advocacy intervention contributed to the established outcomes. For this, we looked in-depth into the contribution of each Alliance to two selected outcomes (or closely related outcome 'clusters'). In addition, we assessed the contributions of the Alliances to other outcomes in more limited ways.

To answer the third evaluation question ('what is the relevance of these changes'), we assessed whether the ILA programmes and outcomes are consistent with the political and public needs and priorities as existing at the various levels: global, interregional, national and local. In answering this evaluation question, we examined whether the programmes under evaluation addressed the needs identified in the Alliances' ToCs and to what extent the changes addressed the Alliances' overall aims for ILA. We also examined whether programmes were relevant for the Southern partners and constituencies.

To answer the fourth evaluation question ('were the efforts of the MFS II Alliances efficient'), we developed an innovative methodology suitable for the specifics of ILA, centring in the 'theory of efficiency' of the ILA programmes. We asked Alliances about their theories of efficiency and endeavoured to evaluate its quality and how they performed against it. The key is that we have shifted emphasis away from an evaluator determining and scoring programme efficiency to establishing how Alliances build in and monitor optimal cost-effectiveness in their programmes.

To answer the fifth evaluation question ('what factors explain the findings drawn from the questions above) we focused analysis on the explanation of the identified outcomes and their relevance. This includes the explanation of contribution, as we only discuss outcomes to which the Alliances claimed to have contributed. In answering this question, we considered internal factors, external factors, their interactions and the nature of the issues addressed.

To obtain the necessary data for our analysis, a range of data collection methods were used. These include in-depth interviews with Alliance members, partners and external experts; analysis of documents including internal Alliance documents and external documents; in-depth case-studies and (in one case) observation. Data gathering mainly took place in 2012 and 2014, while all through the evaluated period evaluators monitored programmes and gathered data on key developments as they took place, when feasible. Because of the complexity of the processes to be understood and the influences of multiple actors and factors that had to be synthesised to create that understanding, the data collected, and consequently the analytical methods used, were primarily qualitative.

Answers to the five evaluation questions

In this summary, we provide answers to the five evaluation questions at a relatively high level of abstraction. For answers including further concretizations, we refer to chapter 17 (Conclusions, lessons learnt and recommendations).

Outcomes

For all Alliances, many outcomes have come in clusters, crossing priority result areas. We find that almost all Alliances, to a larger or smaller degree, have contributed to changes in transnational policy processes around key development issues such as VGGT, UNSCR 1325, REDD+, EU Biofuels policy, RSPO, RTRS and the Post-2015 sustainable development goals. Alliances have thereby inserted civil society voices into ongoing policymaking, often crossing national/international levels, crossing civil society/state/private actor target types, and often

also crossing themes such as environment/social justice, or peace/development. Such clusters of outcomes have typically included a range of different types of outcomes. They have often included organizing CSO collaboration relating to such processes and the provision of input into processes. In some cases, such clusters (also) included increased attention to certain dimensions of issues the uptake of positions by targets, and the incorporation of positions or recommendations into e.g. policy drafting, adjustments of plans, and company behaviours. To clarify their nature and significance, some key examples can be mentioned:

The UN Committee on World Food Security's adoption of the VGGT (Guidelines on the Responsible Governance of Tenure of Land, Fisheries and Forests in the Context of National Food Security) and the subsequent endorsement thereof by other actors like the World Bank and the Dutch government (Impact Alliance).

The Communities of Change Alliance contributed to a cluster of outcomes around policy dialogues on financing UNSCR 1325 on international level, now developing in more concrete discussions to set-up a global discussion group on financing UNSCR 11325 together with UNWOMEN.

Concerning the Renewable Energy Directive of the European Union: the Fair Green and Global Alliance contributed to changed policy in which the increase of the allowed mix of biofuels in fuels for transport is seriously limited now, while reporting on the way the biofuels are produced has become a serious policy issue. Targets that changed their position included the European Parliament, the European Council, the European Commission and the Dutch Government.

ACPF contributed to the policy dialogue on inter-country adoption, influencing national policy processes across Africa and also beyond (Together4Change Alliance).

Also with other forms of transnational institution such clusters of outcome have been attained, as with the Roundtable on Sustainable Palm Oil: a set of outcomes including, among other things, its adopting of a policy for outreach to local NGOs and CBOs and its adoption of a dispute settlement facility to resolve community and company disputes (Ecosystem Alliance).

A few major clusters of outcomes were achieved primarily at (mostly Dutch) national level, as with the reaching of agreement on sustainability criteria for solid biomass among Dutch government, industry and NGOs (Ecosystem Alliance) or the continuation of support for Sexual and Reproductive Health and Rights by the Dutch government, in terms of priority and budget (Sexual and Reproductive Health and Rights Alliance).

Not all clusters of outcomes pertained to existing policymaking processes. Three Alliances achieved major clusters of outcomes on issues through (partly) strategizing relatively

independent from such processes. Impact Alliance has contributed to public awareness, issue uptake, adoption of positions and policy change pertaining to large food and beverage companies' behaviour, with regard to different themes including e.g. gender, land and small-scale food producers by its innovative 'Behind the Brands' campaign. ACPF, part of the Together4Change Alliance, achieved a range of agenda setting outcomes and also some policy changes concerning child rights in Africa, with the AU, UN, and African states, Inter-Agency Working Groups and CSOs, on the basis of its research, reports, expertise and reputation. GPPAC, of the Freedom from Fear Alliance, advanced the development of networks connecting CSOs and a range of other actors including states, RIGOs and international institutions to advance more inclusive and people-centred conflict prevention and peacebuilding.

While clusters of interconnected outcomes have been a prominent pattern, clearly found with, in particular, Fair Green and Global Alliance and Impact Alliance, a few Alliances have more predominantly achieved outcomes that form relatively smaller clusters of outcomes, and in some cases do not (yet) as clearly connect to a larger policy process, even though the themes addressed may be clearly related. This then may be at least partly the result of Alliance partners being involved in different sub-programmes simultaneously, and/or the geographic spread of activities, or of certain (sub) programme being relatively limited in terms of size or success (Ecosystem Alliance, Hivos Alliance, Freedom From Fear Alliance).

The processes of change Alliances are involved in are long-term and highly complex. Individual outcomes, part of such clusters of outcomes, are naturally then mostly intermediate in nature, consisting of steps into the desired direction. Agenda setting outcomes like increased awareness of an issue among targets, or enhanced collaboration with targets, important for all Alliances, are often geared at intermediary target groups. In Clusters I and II, for example, Parliamentary questions, asked in the Dutch and European Parliaments, influenced the terms of debates, as when Dutch Parliament asks the government to do research on Investor-State Dispute Settlement system (Fair Green and Global Alliance). In clusters II and III, we see intermediate nature of outcomes also in relation to targets. Especially when collaborations with UN bodies or RIGOs are developed for the influencing of members states, as with the development of Alliance-like relations between Alliance partner GPPAC and the Human Security Unit at the United Nations for the advancement of Human Security within the UN context (Freedom from Fear Alliance). But also more generally, the intermediate nature of many outcomes is characteristic for most outcomes like, for example the outcome that The EU Commissioner for Development strengthened the EU policy regarding access to renewable energy by publishing a Green Paper that, amongst other things, addressed the issue of energy

access (Hivos Alliance), or the outcome that Palm oil and wood pulp players made public commitments to avoid further expansion on peat (Ecosystem Alliance).

For all Alliances, at least some outcomes involve contributing to the articulating and conveying of civil society voices, and these outcomes have been diverse in nature. One such type of outcome concerns the organizing of civil society, including, for example, the facilitation of networked collaborations of civil society organisations. Such outcomes are, to a relatively large degree, geared towards advancement of Southern partner capacities. Outcomes also contribute to the articulation of views, interests and expertise on the nature of problems and solutions from civil society perspectives. Such (clusters of) outcomes have been attained with Alliance members in the lead, in collaboration with Southern partners, or with Southern partners in the lead within their national or regional contexts. Outcomes have also contributed to the building of connections and interaction between civil society and targets. Finally, outcomes have contributed to the organization and facilitation of platforms and other forms of mutual engagement in collaborative process that open spaces for civil society, facilitating dialogue and more inclusive policy processes in different national and regional contexts. In some cases, these different types of outcomes have contributed to further changes such as policy influence. In many cases though, this is not so (yet), though such influencing is aimed for.

When we consider how outcomes can be allocated to the three priority result areas, we see that they have been mostly achieved in the priority result area of agenda setting. Among many other things, they include outcomes pertaining to the development of advocacy itself, such as the development of strategies and positions amongst Alliance members and partners, and development of relations and collaborations with CSOs in North and South. Engaging with targets resulted in building of relations and collaborations with many amongst these, leading to many outcomes again, including, the gaining of attention for issues, positions and recommendations amongst targets and media; the opening of space for civil society voices in political and policymaking arenas; recognition amongst targets for the offering of valuable contributions to policy; uptake of positions and recommendations in policy process; the influencing of public debate; mobilization of publics. For Cluster III, relatively many outcomes fall in this priority result area.

Also many instances of policy influence can be identified, including demonstrable policy changes, including, among others, the uptake of positions and recommendations in policy or draft policy, changes in accountability structures or normative frameworks, or enhanced commitment to specific policy amongst targets. The degree to which policy influence has been

achieved varies considerably amongst the evaluated programmes. Some programmes were successful at attaining both agenda setting outcomes and policy influence outcomes, whereas some attained mostly agenda setting outcomes. Cluster I, and to a degree Cluster II, contributed to relatively many instances of policy influence.

Fewer outcomes have been achieved in the priority result area of practice change; this is also not what all programmes have focused on. Rather than seeking e.g. the furthering or improvement of policy implementation at lower, 'on the ground' levels, most programmes have focused on policy processes and normative frameworks in national and international institutional arenas. In some cases, programmes did seek changes in practice amongst governments, companies and other actors and, in a number of cases, achieved these.

Outcomes have been achieved with private, state and non-state targets. Most outcomes have been achieved at national levels in North and South, as well international levels. The Dutch government, and in particular the Ministry of Foreign Affairs, has been a target with whom relatively many outcomes were achieved.

When it comes to the way outcomes relate to programmes' overall objectives: Alliances' ILA programmes have sought to advance a range of objectives, and these typically tend to be broadly formulated. A number of outcomes and outcome clusters clearly involve steps towards change in the behaviour of societal actors that may contribute to objectives as formulated. In other cases, contributions of outcomes to objectives are less clear, even when they evidently constitute desired changes, because such objectives are formulated in terms of impact that is relatively at a further remove.

We also see clear differences when it comes to the degree to which programmes have achieved outcomes towards overall objectives. All programmes seek to achieve changes lying in the domains of agenda setting, policy influence and practice change (generally conceived as interconnected and often also as mostly developing sequentially). However, some have mostly achieved outcomes within the domain of agenda setting, whereas others have also achieved policy influence and practice change. This does not necessarily mean that the programmes achieving mostly success in agenda setting are necessarily to be seen as less successful, since they may be operating in more difficult conditions with e.g. relatively much opposition, less attention or lack of opportunity for alignment with important targets. However, such differences do at least indicate relatively larger or smaller progress towards objectives, as conceived in programmes' own Theories of Change.

Relevance

The achieved outcomes have been relevant in light of programmes' Theories of Change. This implies, more generally, that the outcomes contribute to giving shape and influence to civil society on a range of key issues and objectives as selected for support by MFS II. In some cases, relevance of outcomes is very high. In other cases, achieved steps towards objectives were relatively minor, with the ultimate desired changes still far off.

It is not self-evident that achieved outcomes that fit into the Theories of Change are also relevant for the ultimate impact Alliances were aiming for. This impact is outside of the domain of influence of the advocates and the Theory of Change forwards assumptions on the relation between outputs and ultimate effect. Whether these assumptions hold and such relevance indeed exists cannot always be assessed. Achieved changes, also when they can be classified as policy influence, are often at a remove from local realities and implementation, and, importantly, in such cases policy influence often cannot be simply assumed to lead to the impact it seems to promise. This is not to downplay successes; the problem of assessing outcome relevance is directly rooted in the time-span of the evaluation that can be out of synch with the often longer-term advocacy processes Alliances are involved in. Also the complexity and non-linearity of politics and policy process comes in here. Civil society influence is but one among multiple factors in the attainment of often long-term processes of change.

Relevance also needs to be considered here as relevance to constituencies. The programmes varied in the level in which they built on, or involved Southern partners. In some programmes, Southern partners were highly involved; occasionally even leading. In other programmes, there was little involvement of Southern partners; Northern Alliance members in some cases largely acted independently. A higher involvement of Southern partners would not necessarily have led to more relevant outcomes. However, we need to note here that North/South collaborations in a number of programmes did not materialize as proposed or expected.

In the collaboration with Southern partners, attaining the quality and quantity of communication necessary to establish common ground was often difficult. Developing and maintaining commitment to shared objectives, the identification of common interests and the co-creation of activities did not always take place, leading to situations where problems arose in achieving shared ownership, commitment and/or coordination.

Contributions

Outcomes are rarely achieved with Alliances as a sole contributor. Alliances tend to be part of coalitions and networks, sometimes involving many other CSOs. In many cases,

contributions will also have been made by many other types of stakeholders, including e.g. politicians, national governments, international institutions, the public, media and private actors. Furthermore, outcomes showing influence on targets can in some cases be based in already existing alignment or near-alignment in terms of objectives, and results of CSO activities preceding MFS II and Alliance activity more broadly. It would therefore also be not correct to simply equate impressive outcome with impressive contribution.

This being said, we could often establish a plausible contribution to the identified outcomes. In many of the cases we could establish that an ILA programme had a role in the developing and shaping of policy processes and/or changes in policy, but not substantiate the exact magnitude or nature of that role. The evaluation allowed for identifying the qualities and nature of that role, in terms of participation and the nature and role of activities, and the response to these by other actors, including targets as well as other actors.

Programmes contributed mostly through insider strategies, seeking and developing collaborations with targets and partners. In many cases, strategies centred on the contribution of credible content (in the form of e.g. reports, policy briefs, positions, recommendations, testimonies), with this credibility based in quality research and expertise as also the legitimacy and evidence of providing local or constituency voices. In cluster I, outsider strategies were applied. In almost all these cases such strategies were combined with constructive engagement either within the programme itself or through cooperation with other NGOs. Alliances in cluster I were also the only ones that involved actors of the private sector in their strategies, either by targeting specific parts of a specific sector, or by targeting multi-stakeholders commodity initiatives, such as the roundtables for palm oil or soya.

Each of the eight ILA programmes in one way or another, contributed to civil society strengthening. All programmes developed or furthered cooperation and partnerships with Southern partners and in all cases, in one way or the other, this contributed to ILA efforts, although as mentioned above the extent of collaboration varied extensively.

The evaluation found that the relation between advocates and their targets at times shifted, and targets sometimes developed into allies. There have been Alliances working with multi-stakeholder platforms. These platforms deliberately bring together key decision-makers with implementers and stakeholders, in order to bring about inclusive policies. Although decision-making power in these platforms is rarely equal, the distinction between insiders and outsiders in these cases becomes blurred.

Capacity development, including funding and organisational and technical support also contributed to outcomes. NGOs and civil society in the South often have their own sphere of

influence, and capacity development then gets additional meaning, not only rendering advocacy more effective, but also capacitating influential actors to become more responsive decision-makers. This falls outside of the gamut of insider/ outsider strategies.

The evaluation concludes that strategies in themselves cannot be judged as more or less effective. ILA is not a one-size fits all endeavour and each context and objective appears to require its specific mix of strategies. Often, strategies work complementary to each other. No matter which strategy is chosen, the success of ILA depends ultimately on the credibility of the lobby and advocacy. Insider, outsider and hybrid strategies were successful because of the credibility of the messages and materials produced and advanced. Credibility lay also in reputation: the perceived added value of organisations or programme staff, rooted in (perceptions of) experience, knowledgeability, expertise and the 'usefulness' to the targets. Credibility could also be attained by representing legitimate civil society views, and through the ability to link different levels of influence: bringing local voices to international tables and, vice versa, equipping local actors with knowledge on international policy. This latter type of credibility of the representativeness of international advocates can be compromised when collaborations with Southern partners become limited or problematic.

Efficiency

All programmes under evaluation have systematic approaches to efficiency and accountability in place as part of their broader operations in which ILA is embedded. Activities for advocacy were accounted for according to these standard procedures. In this evaluation, we have investigated efficiency in relation to decisions and practices specifically pertaining to advocacy.

The 'theory of efficiency' of Alliances can thus be understood as balancing between effectiveness and efficiency, while ensuring the organisation's identity and principles, including the principle of pursuing a type of efficiency that is considered ethical and a way to retain credibility. All organisations are dealing with efficiency, as time and resources are scarce and ambitions are high, and it is thus embedded in their operations and organisational philosophies. It is understood by the organisations that you cannot always be efficient when you want to be effective, and balancing time, quality and resources plays a major role in decision making on tactics and activities.

While we positively conclude that efficiency is well-considered and practiced by the Alliances, there are certain setbacks as a result of the implicit nature of dealing with efficiency in advocacy decisions, i.e. that efficiency choices embedded in decisions on tactics and

activities tend to disappear from accountability relations, and that efficiency considerations are rarely subject to systematic deliberation or systematic evaluation.

Explanatory factors

While many internal factors contribute to the explanation of findings, most importantly attainment of outcomes, we can establish some that stand out. First, the capability to develop, and commit, to a longer-term vision stood out, including also the capability to develop focus, take decisions, plan and translate these into organisational action. A closely related explanatory factor found of importance here is the capability to select and execute strategy effectively, based in a Theory of Change while relating to the context; this comes down to advocacy competence.

A closely related explanatory factor is staff's capability to monitor environments; adapt to or act relating to that environment and changes in it; build and maintain presence and visibility in arenas form factors that contribute to explaining outcomes. This capability involves continuous context analysis, employing a Theory of Change as guiding, adjusting over time; selection of strategies adequate for context, target and moment, and the using momentum to make the most of opportunities; these can be seen as advocacy competencies.

Such relating to environment always also involves targets. When it comes to relating to targets, explanatory factors involving the capability to relate include the capability to identify and engage relevant targets. We also found that the capability to build and maintain relations and collaborations with partner CSOs contributes to explaining success and failure in the development of relations with lobby targets and constituencies and hence to explain outcomes.

Programme staff's capability to adjust to environments and changes in it, were also found to explain outcomes as well as limits to these, in light of stated ambitions. This capability to adapt and renew also had an internal side. In some cases, programmes showed flexibility and a capacity to learn and adjust in the face of experiences; in a few cases, we saw a limited capability to question assumptions in ToCs. In these cases, programme starting points performed the role of a handbook or script by which to work, organising actors and activities, rather than being taken as points of departure for continuous reflection and reconsideration.

Some programmes worked through focused and coordinated action, contributing to their success. However, this same focus may risk implying a limited involvement and accommodation of diversity, especially with regards to Southern partners.

Finally, we identified the internal factor of maturity of networks involved with ILA as well as the maturity of the programme or programme funded through MFS II. Because building relations with CSOs and targets and the development of agreements on objectives and actions,

as well as their execution, can take years to develop, we found that the amount of time that has passed since the initiation of programmes can be an explanatory factor for outcomes. Some results suggest that more mature programmes had an advantage over less mature ones in terms of attaining outcomes. Funding cycles lasting only a few years may in many case work against effectiveness.

Apart from internal factors, external factors contributed to outcomes. It is important to see here that external conditions and dynamics provide openings or present barriers for Alliances that create opportunities to attain results or limit these. The behaviour of targets emerges as an explanatory factor in different ways. Important aspects in this sense include the target's agenda, positions on issues and their power to influence developments around an issue. Such explanatory factors often also related to the dynamics around targets: Personnel turnover within target organisations, power changes among targets (also positive) and the timing of policy processes were sometimes decisive for ILA opportunities.

Across the evaluated programmes, dimensions of context also emerged as explanatory factors. Political space for CSOs in specific geographic contexts impacted the possibilities to undertake activities, affecting the outcomes achieved. The cultural context of activities and of the issue or agenda on which ILA is conducted sometimes similarly influenced the opportunities to undertake activities and find a hearing or establish forms of collaboration. Institutional openness to civil society participation and influence also emerged as a factor. Changed conditions through wars and disasters too impacted the extent and nature of outcomes. Another contextual factor can be identified in the political support for ILA by CS, including budget and budget cuts for ILA.

Finally, the nature and context of the issues that the ILA addressed emerged as an explanatory factor in a range of ways. First of all, the way specific audiences (i.e. targets, publics and partners) relate to issues partly defined opportunities for programmes to achieve outcomes. The construction and resonance of issues within societies and with targets is a factor that explains Alliances' chances of success at attaining outcomes. Emergence of issues involves a complex process that programmes can influence only to a limited degree.

Issues can also be highly sensitive in specific contexts, and this can influence the possibilities for advocacy, as ILA on such issues may be considered controversial and may be countered by other actors including targets as well as other CS actors. Similarly, ILA on certain issues was controversial because it threatened vested interests; such situations had consequences for the space for ILA, also because opposition threatened the security of CS actors.

Finally, the complexity of the issues on which the Alliances conducted ILA may have implicated programmes' chances for achieving results. For example, an issue like 'human security' (the advocacy issue of one evaluated programme), involves multiple actors, understandings, agendas, institutions and levels to a degree that it may make the advancement of objectives much more challenging than in other cases, where understandings, problems and solutions are more clear and shared, and institutional and political contexts are more demarcated and relatively less conflict-ridden.

In conclusion: organisational capacity helps to explain the effectiveness of the evaluated programmes, and we could identify a range of capabilities that came out as important. However, these capabilities are only one factor among several that need to be taken into account while assessing programmes' effectiveness. The nature and contribution of different factors to explanation differ per programme. This implies that it is not possible, on the basis of this evaluation, to identify what 'works' and what 'doesn't work' in a general sense.

Lessons learnt and recommendations

Based on the results of research focused on the answering of the five evaluation questions, the development and application of methodology, as also a range of unexpected learnings that could take place because of our close involvement with the Alliances' work over 2012-2014, the evaluation lead to the identification of a number of 'lessons learnt' that may further ILA effectiveness. It needs to be stressed that these lessons emerge from our study of the work of the Alliances, and are often based on good practices that were observed. The lessons form the foundation for a set of recommendations to funders, development and advocacy professionals and evaluators.

Conceptualizations of advocacy

The classical approach to advocacy implicitly takes advocacy as oriented towards the influencing of decision makers' understandings, views and actions. While this holds to a large degree, this approach does not explicitly conceptualize the realities of present-day governance that evaluated programmes engage with. Such realities include bilateral and multilateral engagements with intergovernmental, governmental, research, CSO, private sector, military and semi-governmental organisations. Combining outsider strategies and insider strategies often lead to hybrid approaches and influencing is done in collaborative structures, where 'target' of advocacy sometimes become 'allies'. A conceptualization of advocacy that explicitly

does justice to these realities is due for ILA. To a large extent, the Alliances in this evaluation consciously work with such a conceptualization, whereas for some Alliances this is more tacit.

Recommendation 1: Funders and evaluators, and to some extent Alliances, need to approach ILA with an eye for the complex realities of governance, where the nature of relations between CSOs, decision-makers and other actors often defies a simple relation of advocates influencing targets, and where engagements and interactions are diverse, dynamic, and often long-term.

Working with Theories of Change

Some programmes employed Theories of Change effectively for articulation, organization, analysis, action, reflection, and adjustment on the base of experience. Such Theories of Change were specific rather than broad, and put to the test in the execution of programmes. This was not always the case, and sometime the Theory of Change figured as a script, rather than being taken as point of departure, for continuous reflection and reconsideration.

Relatedly, advocacy planning, monitoring and evaluation in a few cases turned out to be centred on activity and process rather than result-oriented strategizing. There is a clear tension here between the need to work in a structured fashion, and the ability to adjust to changing circumstances and to learn from experience. A major lesson stemming from the evaluation, is that ILA programmes should use the Theory of Change (or related tools) in an active way throughout the programme: revisiting Theories of Change and employing these as tools for introspection. This evaluation underlines the importance of applying a tool like the Theory of Change for advocacy in such a way that the sphere of influence is clearly demarcated and assumptions regarding necessary interventions and expected outcomes towards the final objective are spelled out.

Recommendation 2: Funders, advocates and evaluators can further learning, and thereby effectiveness, by systematically working with Theories of Change (or similar/related tools) as tools for developing and adjusting understandings and strategies.

Recommendation 3: Funders and advocates and evaluators can work with Theories of Change (or similar/related tools) for making explicit and assessing programmes' (potential) relevance in the light of constituency needs and priorities.

Assessment of outcomes

In this evaluation, we have sought to uphold a nuanced understanding of effectiveness that does justice to the complexity of advocacy. Advocacy is often to be understood as long-term

investment, with outcomes mostly consisting of steps towards desired outcomes, with agenda setting outcomes often understood as a step towards achieving policy influence or change in practice. Cases where policies or practices have demonstrably changed are much fewer. In some cases influence may be seen as possibly taking more time, with 'failure' to attain influence not automatically needing to be seen as ineffectiveness, considering the time needed for change and the fact that maintaining an issue in view can already be an achievement. With ILA, there is often a time to sow and a time to harvest, and this in unpredictable ways.

It is important for advocates to incorporate such process indicators in their monitoring, and avoid overly emphasizing results. Funding and evaluation cycles need to move away from a linear input-output-outcome approach to assessing effectiveness, and towards an approach that focuses on the qualities of programmes and processes and the ways they relate to the complex challenges of ILA. Internal and external monitoring systems and assessment needs to put more emphasis on organisations' capacity to analyse where strengths and opportunities lie for their programmes and the qualities of their acting on such analyses.

Recommendation 4: Funders, advocates and evaluators should not establish effectiveness on the basis of achieved outcomes alone, but also on the merits of advocates' way of strategizing in the face of the complex challenges of the environment they operate in and seek to change.

For the assessment of Alliances' contribution to outcomes, the evaluation team has done in-depth analysis of a limited set of outcomes. This was resource-intensive and evaluators could consider to aim for more overview-like analyses of contribution. While establishing contribution was often feasible in case of the priority result areas of agenda-setting, it was less doable for policy change or change in practice, especially in international arenas involving multiple CSOs, targets and institutional levels.

Recommendation 5: Be realistic about the feasibility of contribution analysis for different types of outcome and adapt methods for contribution analysis accordingly.

How to enhance effectiveness

The evaluation concludes that strategies in themselves cannot be judged as more or less effective. ILA is not a one-size fits all endeavour and each context and objective appears to require its specific mix of strategies. Often, strategies work complementary to each other. Organizational capacities, external factors and issue-related factors, all in diverse ways, contribute to explanations of success and failure. Success therefore cannot be straightforwardly equated with effectiveness, and failure can happen in spite of high organizational capacities.

The evaluation thus points to the need to consider different factors explaining effectiveness, in their interrelatedness, when assessing advocates and programmes.

Recommendation 6: To further effectiveness, funders, advocates and evaluators need to consider internal, external and 'issue-related' factors together. By this they can identify organizational and programme strengths and weaknesses, see where challenges lie and learn how to handle these.

Connecting 'global' and 'local'

Alliances commonly invested in relations with civil society in the South. In some cases, choices of partners were based on the added value expected from specific actors, while in a few cases, considerations of inclusiveness and openness were more central to the development of networks. Collaborations with Southern partners contributed to outcomes in important ways, and Alliances sought and often succeeded in developing relations, outputs and outcomes effective for the different actors involved. But this also turned out challenging. We sometimes found Southern partners experienced a 'disconnect' between Alliance efforts and local realities or priorities.

While relations and collaborations with Southern partners have been important for almost all programmes, we note that communication, collaboration and accountability structures in relation with Southern partners and constituencies have not been explicit issues for ILA under MFS II, or its evaluation. While this does not imply that CFAs and partners were not accountable to each other, this is a lacuna. Working with 'partners' as such is no guarantee for the constitution of a voice that reflects constituency views and priorities.

Recommendation 7: To further inclusiveness of ILA outcomes, funders, advocates and evaluators need to make communication, collaboration and accountability structures in relation with Southern partners and constituencies integral to development, execution, monitoring and evaluation of ILA programmes.

Ownership and effectiveness

The networked nature of advocacy often added to effectiveness, and the diversity amongst Alliance members and partners was often fruitful. However, the diversity of networks sometimes also brought about problems. The most important of these were focus and ownership – with these two often being in tension with each other. Effectiveness and ownership can be in tension with each other, with potentially important trade-offs between these leading to sacrifices

on either front in networks bringing together diverse actors. Balancing effectiveness and ownership emerged as a key tension in this evaluation.

Recommendation 8: Funders, advocates and evaluators of ILA need to explicitly address potential tensions between effectiveness and ownership in the development, execution and evaluation of partnerships and programmes.

Time-frames and effectiveness

Achieving voice for civil society takes time. In some Alliances, depending, for example, on the 'age' of the issue and the existence of potential partners and coalitions at the start of the MFS II period, achieving outcomes such as credibility and recognition took several years. Only after having achieved recognition will invitations for further collaboration and expressions of interest come.

Recommendation 9: To further effectiveness of programmes, donors and advocates need to consider the time frames required to develop a credible voice, recognition and relations. As time frames often exceed project durations, this asks for medium- to long term visions and funding strategies.

Reflexive monitoring of efficiency

All programmes under evaluation have systematic approaches to efficiency and accountability in place as part of their broader operations in which ILA is embedded. Activities for advocacy were accounted for according to these standard procedures. Our focus in the evaluation was how efficiency played a role in decisions and practices specifically pertaining to advocacy. While positively concluding that efficiency is well-considered and practiced by Alliances, we also observe that these considerations are not subject to systematic deliberation, or systematic evaluation. They are also not brought into accountability relations. With resources scarce and with effectiveness holding no direct and self-evident relation to spending, result-orientation as much as accountability across partnerships would be well served by inclusive and reflexive monitoring of efficiency choices and their outcomes.

Recommendation 10: The field of ILA would be well served by the development and use of inclusive ways to monitor, reflect on and account for efficiency choices and their outcomes.

Evaluation of programmes

This evaluation took up 8 very different programmes, and was carried out by a team of 12 evaluators. While the diversity allowed the emergence of a number of cross-cutting insights,

the requirements of consistency and coherence across evaluated programmes was very challenging and time-consuming. The grouping of the 8 Alliances in three Clusters was useful as a structure for the evaluation team, but the requirement to submit Cluster-level assessments was unproductive, in the view of the evaluation team.

Recommendation 11: For (possible) future joint evaluations, devise a structure that does more justice to diversity and seeks to develop cross-cutting analyses in more productive ways.

Evaluation orientation

The five evaluation questions and evaluation structure devised for this evaluation focus on assessment of effectiveness with most analysis taking place at the start and end of the funding cycle, which was not conducive to learning during the programmes. In hindsight, the evaluation team and the Alliances could perhaps have been more pro-active in incorporating a learning agenda. Nonetheless, inclusion of more learning oriented evaluation questions, and a more learning-oriented process structure, is advisable for future evaluations. Such a setup might also facilitate trust-building better than the present setup.

Recommendation 12: To facilitate learning, evaluation design needs to be better tailored to learning ambitions.

Annex 2

Overview of the interviews and (in-) formal meetings 2012-2015

Interviews and (in-) formal meetings (see: Arensman et al, 2015)

Type of interviews and meetings	Alliances under evaluation 2012-2015	# of interviews and (in-) formal meetings 2012-2015
Staff of the alliances under evaluation, partners, and advocacy targets	Together 4 Change	80
	Freedom from Fear	80
	Communities of Change	35

Interviews and meetings held between 2012-2015

Interviewees	# of interviews per program	Total #
Evaluated program staff, including advocates, monitoring and evaluation officers, program officers and policy officers (including board members).	ACPF: 24 GPPAC: 41 WLPS: 11	76
Policy makers in governments, regional-intergovernmental organisations (AU, EU, ASEAN), international institutions (including advocacy targets).	ACPF: 12 GPPAC: 5 WLPS: 2	19
Civil society stakeholders including NGOs, partners within networks and social movements (direct and indirect to the evaluated programs), including Inter-Governmental organisations (UN).	ACPF: 44 GPPAC: 34 WLPS: 22	100
Total # of interviews in the MFS II ILA Evaluation	Other programs: 209	404
Meetings with the evaluation team		15
Meetings between evaluation team and the evaluated organisations, including meetings with the Netherlands Ministry of Foreign Affairs.		29

Annex 3

Dutch Summary

Negotiating Effectiveness: The politics of Results in advocacy for development and its evaluation

Bodille Arensman

In deze doctoraal dissertatie analyseer ik hoe in de praktijk van advocacy binnen de internationale ontwikkelingssamenwerking betekenis wordt gegeven aan effectiviteit en de evaluatie van effectiviteit. Hiernaast analyseer ik hoe de evaluatie van effectiviteit beïnvloed wordt door een resultatenpolitiek (*politics of results*). In de verschillende hoofdstukken presenteer ik mijn bevindingen als antwoord op de vraag: *hoe wordt in de praktijk van advocacy binnen de internationale ontwikkelingssamenwerking betekenis gegeven aan effectiviteit en de evaluatie hiervan en hoe speelt resultatenpolitiek hierbij een rol?*

Beïnvloedingsprocessen zoals advocacy spelen een steeds belangrijkere rol als strategie voor effectiviteit in het nastreven van structurele veranderingen binnen de ontwikkelingssamenwerking (Hudson, 2002; Jolly et al., 2004; Jones, 2011; Keck and Sikkink, 1998; Kremer et al., 2009; Teles and Schmitt, 2011). Deze veranderingen in de ontwikkelingssamenwerking kosten vaak veel tijd en een lange adem en er zijn vaak veel verschillende actoren betrokken. Deze actoren werken en opereren in verschillende contexten en op verschillende niveaus binnen diverse organisaties, veelal over landsgrenzen heen. Om hier nog een schepje bovenop te doen gaat advocacy vaak over ongrijpbare en niet zichtbare resultaten, terwijl de druk op het aantonen van effectiviteit juist meetbare en zichtbare resultaten vraagt. Veel van wat er gebeurd binnen advocacy is strategisch van aard. Advocates handelen en reageren op mogelijkheden en veranderingen in veranderende contexten. Ze onderhandelen over en manoeuvreren tussen verschillende belangen, behoeften en vragen, in interactie met verschillende actoren. Evaluatoren en advocates werken dus veelal in contexten waarin kennis over advocacy effectiviteit en praktijk vaak impliciet zijn. Dit is een reden dat er binnen het veld van advocacy en advocacy effectiviteit geen gezamenlijk perspectief is op ‘wat werkt, hoe en wanneer.’ Het vaststellen van advocacy resultaten en de effectiviteit van advocacy inspanningen is daardoor erg ingewikkeld, maar bovenal ook politiek gevoelig.

Ondertussen is er steeds meer nadruk vanuit het publieke en politieke domein op meetbare en zichtbare effectiviteit. Dit is zichtbaar aan de hand van de ontwikkeling van modellen en methoden die een voorspelbare effectiviteit nastreven en zich daarbij richten op zichtbare resultaten, liefst meetbaar binnen de bestaande financieringscycli. Deze ontwikkelingen beïnvloeden de manieren waarop organisaties werken door middel van de procedures en protocollen waarmee bepaalde doelen worden nagestreefd. Er is hierbinnen een resultatenpolitiek of *politics of results*. Dit is het politieke en strategische proces rondom resultaten, waarbinnen onderhandeling plaats vindt over de betekenis van effectiviteit. Met oog op deze ontwikkelingen laten de bevindingen in deze dissertatie zien dat advocacy effectiviteit en evaluatie beïnvloed worden door deze *politics of results* en als zodanig sociaal en politiek geconstrueerd zijn.

In deze dissertatie laat ik zien dat advocacy effectiviteit moeilijk vast te stellen is en tevens ingewikkeld is om te operationaliseren. Dit leidt ertoe dat effectiviteit meestal wordt afgemeten aan de vraag of de organisatie haar van te voren vastgestelde doelen heeft behaald (*politics of results*). Hieraan wordt vervolgens betekenis ontleend of de organisatie in de ogen van de stakeholders legitiem en geloofwaardig is. Hierbij gaat effectiviteit dus over of een interventie haar doelen behaald in tegenstelling tot hoe deze doelen worden behaald. Echter, de bepaling van effectiviteit en hoe bepaalde acties leiden tot resultaten en meer of minder effectiviteit is niet vanzelfsprekend. De hoofdstukken in deze dissertatie laten dan ook de betekenisgeving aan effectiviteit zien als een menselijk proces, waarin interactie leidend is en intrinsiek politiek gemotiveerd is.

Het onderzoek voor deze dissertatie naar advocacy binnen de ontwikkelingssamenwerking en de evaluatie hiervan, was onderdeel van één van de grootste advocacy evaluaties wereldwijd (MFS II ILA Evaluatie). De evaluatie vond plaats van 2012-2015 met twee meet momenten (2012 en 2014), waardoor het meer-jaren programma ruimte bood om er een PhD in te verweven – met toestemming van de donor en alle betrokken actoren. De duale rol van evaluator en PhD onderzoeker, bood een unieke mogelijkheid om de processen en praktijken van advocacy en haar evaluatie te observeren, bestuderen en analyseren. In het onderzoek richt ik mij op vier case-studies. Drie case-studies zijn de geëvalueerde advocacy programma's en hun evaluatie. De vierde case-study is de evaluatie als proces op zichzelf. Tijdens de evaluatie werd bijvoorbeeld duidelijk dat er veel politieke belangen een rol speelden. Deze belangen kwamen naar voren in de wisselwerking tussen evaluatoren en geëvalueerden. Dit werd duidelijk in de interacties over conceptualisering van effectiviteit; in de interacties gericht op het identificeren, beoordelen en presenteren van resultaten; en in de strategische

interacties en onderhandelingen als onderdeel van de evaluatie. Vooral rondom resultaten werden veel belangen zichtbaar, zoals over de toegang tot informatie of hoe resultaten werden geïdentificeerd en gepresenteerd. Gedurende het proces van de evaluatie heb ik geleerd dat 'het eerlijke verhaal' over effectiviteit veel meer gelaagd en multidimensionaal is dan ik aanvankelijk dacht.

Hoofdstuk één is een inleidend hoofdstuk over advocacy in ontwikkelingssamenwerking en de evaluatie ervan. In dit hoofdstuk worden de theoretische discussies uiteen gezet, zoals de conceptualisering van de resultatenpolitiek (*politics of results*), de betekenis van effectiviteit en de tradities van evaluatie. Hoewel de literatuur belangrijke inzichten laat zien, is er weinig aandacht voor advocacy effectiviteit en hoe deze betekenis krijgt of hoe het beïnvloed wordt door de politiek rondom resultaten, met betrekking tot verantwoording, geloofwaardigheid en legitimiteit van de organisatie. Terwijl advocacy steeds meer wordt ingezet als strategie in de ontwikkelingssamenwerking, is er vooral over de specifieke uitdagingen rondom advocacy effectiviteit en de meetbaarheid hiervan een gat. Het dynamische karakter van advocacy en de gelaagde contexten waarbinnen advocacy evaluatoren opereren (zoals de *politics of results*)maakt dat de evaluatie van effectiviteit ingewikkeld is. Dit alles maakt de vraag relevant hoe effectiviteit betekenis krijgt in de context van dergelijke processen.

Hoofdstuk twee analyseert het spanningsveld tussen het nastreven van eigenaarschap (ownership) en effectiviteit binnen een transnationaal advocacy netwerk. Eigenaarschap en effectiviteit zijn beide politieke doelen binnen de ontwikkelingssamenwerking en beide zijn vaak ingebed in donor overeenkomsten als onderdeel van de financieringsrelatie met organisaties. De twee concepten dagen elkaar uit terwijl ze tegelijkertijd bijdragen aan elkaars betekenisgeving. Effectiviteit wordt veelal gevraagd in meetbare, zichtbare termen terwijl eigenaarschap zich hier bij uitstek niet goed voor leent. De *politics of results* in ontwikkelingssamenwerking stelt een directe relatie tussen eigenaarschap en effectiviteit. Echter, het nastreven van eigenaarschap vraagt om een andere invulling van effectiviteit, namelijk als ambigue, multidimensionaal concept dat betekenis verkrijgt door sociale en politieke interacties en afwegingen. De bevindingen in dit hoofdstuk laten zien dat niet zomaar kan worden aangenomen dat eigenaarschap ook tot effectiviteit leidt en dat de manier waarop effectiviteit betekenis krijgt van belang is in het begrijpen ervan. De relatie tussen de twee concepten van effectiviteit en eigenaarschap is ambigue. Dit hoofdstuk analyseert deze relatie in drie dimensies: de collectieve identiteit, het verantwoordingsproces en de advocacy boodschap. De vraag gaat hierbij niet over het meer of minder effectief zijn, maar over het

proces dat betekenis geeft aan effectiviteit in het bijzonder in specifieke institutionele constellaties en interacties.

Hoofdstuk drie analyseert hoe effectiviteit onderhandeld wordt binnen advocacy evaluatie. Door de *politics of results* worden resultaten als de kern van effectiviteit gezien. Hieronder gaat de aanname schuil dat resultaten objectief gemeten, geïdentificeerd en gepresenteerd kunnen worden. De literatuur gaat hier echter maar in beperkte mate op in hoe dit beïnvloed wat er wordt geëvalueerd, hoe en waarom. In dit derde hoofdstuk beargumenteer ik dat er een verschuiving nodig is weg van de huidige positivistische trends die zich richten op meetbaarheid, resultaten en result-based management. Een dergelijke verschuiving is met name nodig omdat advocacy effectiviteit zich niet laat vangen in deze nauwe termen van meetbaarheid wat vraagt om resultaten die op korte termijn zichtbaar zijn. Advocacy evaluatie zoals dit hoofdstuk aangeeft is juist bij uitstek subjectief. Deze subjectiviteit zorgt voor interessante mogelijkheden tot het reflecteren op en ontwikkelen en aftasten van nieuwe inzichten in de waarde en betekenis van advocacy processen. In dit hoofdstuk geef ik een aanzet tot het creëren van ruimte om tot nieuwe inzichten te komen in de diverse realiteiten van evaluatie door in te zoomen op de belangen die een rol spelen voor zowel evaluatoren om een kritische assessment te maken als de geëvalueerden die het niet altijd eens zijn met de gestelde assessments.

Hoofdstuk vier analyseert de veel gebruikte methode - Theory of Change – en stelt de vraag of deze wel voldoet aan de verwachtingen dat het een methode is die bij uitstek geschikt is voor complexe en dynamische interventies zoals advocacy. Terwijl Theory of Change veelvuldig gebruikt en geprezen wordt door ontwikkelingsorganisaties, evaluatoren, donoren en beleidsmakers, stel ik in dit hoofdstuk dat de methode tekortschiet in de manier waarop het nu geïmplementeerd wordt. Theory of Change als evaluatie methode wordt momenteel nog te veel vast gezet in theoretische aannames, waarbij weinig ruimte is voor onvoorspelbare, niet lineaire en dynamische processen, zoals advocacy. Dynamische processen zoals advocacy vragen om meer inzicht in de theorie in relatie tot de praktijk, dus hoe advocates strategisch handelen, acteren, aanpassen en manoeuvreren in de dagelijkse praktijk. In het nastreven van structurele veranderingen handelen advocates namelijk strategisch terwijl ze verschillende en vaak veranderende contexten, systemen en organisaties navigeren en zich hierin manoeuvreren. De resultaten van deze acties zijn dan ook niet altijd duidelijk of zichtbaar. Bestaand onderzoek geeft slechts heel beperkt inzicht in wat dit betekent voor advocacy en de evaluatie hiervan.

In hoofdstuk vier analyseer ik daarom hoe advocacy betekenis krijgt door middel van het strategisch handelen van de advocates waarbij resultaten vooral voorkomen uit interacties en

niet altijd gepland kunnen worden, maar eerder opkomen in het moment waarin mogelijkheden gecreëerd worden. Hoofdstuk vier stelt voor om de Theory of Change te versterken door een betere inbedding in de praktijk van advocacy: *twinning theory of change with practices of change*. Hiermee kan een brug geslagen worden tussen de theorie van evaluatie methoden en de strategie als praktisch handelen en iets dat actoren doen in plaats van als iets dat organisaties hebben. Hierdoor wordt tegelijkertijd ruimte gecreëerd om juist de advocates het centrale middelpunt van de evaluatie te laten zijn. Advocates geven namelijk interpretatie en betekenis aan de effectiviteit door middel van hun praktische inzichten, strategisch handelen en manoeuvrerend vermogen in veranderende omgevingen in interactie met andere actoren.

Als verdieping op de vorige hoofdstukken, analyseer ik in hoofdstuk vijf het problematische karakter van advocacy resultaten die vaak niet zichtbaar of meetbaar zijn. De identificatie van resultaten is afhankelijk van diverse interpretaties van actoren die erbij betrokken zijn. Hierdoor kunnen advocacy resultaten niet gezien worden als vanzelfsprekend. Veel gebruikte methoden voor de evaluatie van advocacy resultaten zijn Outcome mapping en Outcome harvesting. Zij bieden echter weinig richtlijnen wanneer het gaat om de problematische karakter van de identificatie van deze resultaten. Dit hoofdstuk laat zien dat de resultaten hiervan politiek zijn in termen van hoe ze worden geïdentificeerd (geclaimd of juist niet geclaimd, over-reported of juist under-reported) om aan de doelen van de evaluatie tegemoet te komen (het zichtbaar en meetbaar maken van resultaten). In andere gevallen zijn de identificatie van resultaten en het beoordelen van effectiviteit juist onderhevig aan interpretaties (diverse realiteiten), geconstrueerd (multi-gelaagd) en politiek (multidimensionaal, diverse belangen spelen een rol) in plaats van objectief meetbaar. In het geval van advocacy moeten resultaten daarom geconstrueerd en gereconstrueerd worden. Dit betekent dat processen, praktijken en resultaten bevraagd en geïnterpreteerd moeten worden op een kritische, nieuwsgierige en open manier. Evaluatoren moeten hierbij meer aandacht geven aan het proces en de praktijk waarin advocacy effectiviteit haar betekenis krijgt en waarin resultaten vorm krijgen.

Het laatste hoofdstuk is het concluderende hoofdstuk waarin ik de bevindingen van de voorgaande hoofdstukken bij elkaar breng en een overkoepelende analyse presenteer. Tevens zet ik in dit hoofdstuk uiteen hoe de bevindingen bijdragen aan een toekomstige onderzoek agenda en agenda voor de praktijk van evaluatoren, advocates en beleidsmakers. Ten eerste is er in de evaluatie van advocacy sprake van formele en informele betekenisgeving rondom effectiviteit. Informeel werd er ruimte gecreëerd voor discussies en gezamenlijk leren. Dit bood inzichten in fundamentele dynamieken, uitdagingen en contexten waarin afwegingen, onderhandelingen en interpretaties richting gaven aan strategische besluitvorming. Formeel

werd er veel belang gehecht aan de waarde van resultaten. De geëvalueerde organisaties wilden graag hun resultaten zichtbaar maken als successen (over-reporting, claiming, zichtbaarheid), vooral in de geschreven rapporten. Ten tweede speelde in deze processen de *politics of results* een belangrijke rol in het nastreven, onderhandelen, begrijpen en identificeren van effectiviteit (multidimensionaal). Er speelden belangen een rol in de uitdagingen rondom behoeften (legitimiteit, geloofwaardigheid) en vraag (verantwoording) bij zowel evaluatoren als bij de geëvalueerde. Ten derde stel ik daarom dat de evaluatie van advocacy een andere manier van evalueren nodig heeft. De huidige evaluatie methoden doen nog onvoldoende recht aan of bieden onvoldoende ondersteuning voor het begrijpen en evalueren van dynamische en complexe processen zoals advocacy. Juist het dynamische karakter vraagt om een evaluatie benadering met een focus op betekenisgeving met als centraal uitgangspunt het strategisch handelen van actoren en de strategie als praktijk. Concluderend beargumenteer ik dat er een verandering nodig is als het gaat om het denken over effectiviteit in de ontwikkelingssamenwerking. Een verandering uit de comfort-zone van benaderingen en methoden die vooral gericht zijn op controleerbaarheid, voorspelbaarheid, meetbaarheid en zichtbaarheid. Er is juist behoefte aan een verandering richting de betekenis van praktijken en processen inclusief interacties, strategisch handelen, besluitvorming, organisatie structuren, rollen, waarde, en onderliggende belangen en afwegingen.

Bodille Arensman

Wageningen School of Social Sciences (WASS)

Completed Training and Supervision Plan

Name of the learning activity	Department/Institute	Year	ECTS*
A) Project related competences			
Workshop and evaluation training	MFS II International Lobbying and Advocacy evaluation	2012	4.5
Workshops, interviews and feedback meetings, presentations and discussions with Dutch co-financing agencies in the Netherlands and abroad	MFS II International lobbying and Advocacy Evaluation	2012-2015	5
Learning day organization, presentation, facilitation and interaction with co-financing agencies and Partos	MFS II International lobbying and advocacy evaluation learning day	2014	3
Evaluation methodology development	MFS II International lobbying and advocacy evaluation	2012-2015	5
B) General research related competences			
Introduction course	WASS	2012	1
Qualitative writing and data analysis	WGS	2015	1.5
Scientific writing course	WGS	2014	1.8
Presenting with Impact	WGS	2015	1
Issues in political, policy and organisation Ethnography	European Consortium for Political Research (ECPR) Summer School in Methods and Techniques	2013	2
"The politics of results in advocacy evaluation: the case of a child rights advocacy organisation"	Conference in Brighton by Institute of Development Studies (IDS)	2013	1
Proposal writing	WUR	2012-2013	6
"Challenges in evaluating advocacy for development"	Netherlands Institute of Governance Conference	2014	1
"Negotiating effectiveness in advocacy evaluation"	American Evaluation Association	2015	1
"Negotiating effectiveness: the case of a transnational advocacy evaluation"	Development Studies Association	2016	1
"Unintended outcomes: can ToC make a difference? The case of a transnational advocacy evaluation"	Netherlands Ministry of Foreign Affairs	2017	1
C) Career related competences/personal development			

Training on advocacy evaluation for practitioners	NEDWORC.	2014	2
International lobbying and advocacy workshop	Netherlands Ministry of Foreign Affairs	2014	2
Spindle Awards summer labs: "Narrative Assessment"	Partos Spindle Awards	2017	4
Total			**43.8**

*One credit according to ECTS is on average equivalent to 28 hours of study load

Colofon

Financial support by: NWO WOTRO (WO7.72.601)

Cover design & painting by: Elizabeth-Anna van Putten

English editing by: Dr. Jennifer Barrett

Printed by: DigiForce || ProefschriftMaken