Negotiating effectiveness in transnational advocacy evaluation

Bodille Arensman
Wageningen University & Research, the Netherlands

Margit van Wessel
Wageningen University & Research, the Netherlands

Abstract
International development organizations 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 article 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 article 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, effectiveness, international development, negotiation, outcomes

Introduction
In the last two decades, non-governmental organizations (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, which is assumed to be objective and independent, to demonstrate legitimacy and account for funding streams (Eyben et al., 2015; Riddell, 2014). A deeper understanding of advocacy evaluation is needed, because evaluating these efforts and their influence demands different evaluation approaches.

This article makes three central arguments. First, advocacy is extremely difficult to evaluate. In this context, the relation between funds and achievements is ambiguous (Coffman, 2009; Coffman and Reed, 2009; 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 illustrate how effectiveness was necessarily identified through direct interaction and negotiations with staff members of the evaluated organizations. These organizations were highly interested in favourable results, so establishing effectiveness was of major concern. This article 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.
Advocacy challenges realist evaluation

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 (Gulrajani, 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 (Gulrajani, 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 et al., 2010; Eyben et al., 2015). These demands inspired theory-based and objective-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 example, 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 characterized by a search for the truth. However, not everything can be measured in the hard sense of the word. Constructionists emphasize 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 generalizable 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 organizational and professional survival (Eyben et al., 2015). Approaching evaluation from an objective perspective does not allow for challenges to these power relations, as it assumes evaluation roles,
evaluation 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.

**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 emphasizes 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 (McDonald, 2008; Markievicz, 2005; 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 emphasizing the negotiating and facilitating roles of the evaluator (McDonald, 2008; Markievicz, 2005), 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.
Methods and case

The evaluation used as a case study in this article 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 organizations. The evaluation was coordinated by the Netherlands Organization 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 article 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 article 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 organizations and civil society) and other relevant stakeholders in the field (Arensman et al., 2015). Where possible, participant observation was also conducted.

In this article, we focus mainly on the three advocacy programmes the authors were closely involved in evaluating. We followed the diverse advocacy programmes through time (2012–15) and space (diverse geographic locations and political arenas worldwide), conducting a multi-sited ethnographic field study (see Marcus, 1995). The analysis for this article 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.

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 among 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 conceptualization 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 organized 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.

**Shaping the evaluation process and outcomes**

*Identifying outcomes*

Conceptually and practically, organization 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 organized, 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 organization. Although this provided meaningful insight into the wide diversity of projects the organization 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 recognized 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 specialized 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 organization 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 organization 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 organization 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 organization 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’ characterized this organization’s way of working, and they were applauded for this by the stakeholders and policy makers with whom they worked: ‘[The organization] 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 organization was able to open space for discussions on these issues precisely because of their high regard for these sensitivities. Thus, the under-reporting actually contributed to the advocacy outcomes.

This organization’s reluctance to claim influence or changes was understandable, but it left the organization unaware of the extent of their successes, and it made their achievements invisible. In discussions with the evaluators, the Dutch partner organization emphasized 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 organizations or ad hoc cooperation efforts, whereas others were more permanent organizational 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 conceptualizations 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 conceptualizations touched upon the core of organizations’ 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 organizations considered networking a strategy rather than an outcome; to them, it was an integral part of advocacy in a globalizing 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 organizations.

**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 were measurable. However, it did not include conceptualizations 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 organizations 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 organizations 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’ (Evaluation manager, 2013), emphasizing 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 emphasized 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 organizations. Our report of this assessment could be more or less positive, potentially with significant consequences for the evaluated organizations. Therefore, handling our search for information continued to involve strategic choices by organizations 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 organizations.

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 organization 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 stylized and idealized 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 organization 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 organizations 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.

**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 finalized. 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 1 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 emphasized the need to at least ‘interview all relevant people’. The team’s work was criticized 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 organizational 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

| 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** |
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 organization, it was important to
emphasize 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 organization] also plays a role as convenor.

The organization’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 organization’s
claims, as was the case for ‘direct lobby’. We found that relations were built with policy mak-
ers, 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 organizations. 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 organizations’ feedback on our evaluation report tended to add to and reframe the find-
ings 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 emphasize the pro-
gramme’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. (Programme 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 organizations 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., 2016). 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 organization’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 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 emphasized 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 emphasize the layered nature of being a network. We emphasized the process and included networking-specific outcomes. We also incorporated the organization’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.

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.

Conclusions

Advocacy influences people’s daily lives, as it is increasingly pursued as a strategy by development organizations, 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 article 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.

Acknowledgements

The researchers are grateful for the openness and cooperation of the staff of the evaluated programmes, members and partners, and for the opportunity provided by the evaluation team to study an advocacy evaluation and reflect on the processes and findings. Special thanks to Dr Jennifer Barrett for many meaningful suggestions and language editing.

Funding

This work was part of a PhD research project supported by the Netherlands Organisation for Scientific Research.
Notes
1. These are often policy and/or impact evaluations.
2. Widely used evaluation guidelines see for example IOB (2009); OECD (1991).
3. These methods are often in line with realist evaluation (see Pawson and Tilly, 1997), focusing on what works and gathering information on causal mechanisms. Policy evaluation has a theory-based tradition that seeks to determine what was achieved following the policy and with the funding spent.
4. Negotiation is understood as socially and politically constructed, following Putnam (2004).
5. For further information, see Arensman et al. (2015).
6. For a full list, see Arensman et al. (2015: 672).

References
Alexander J, Brudney JL 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): 565–70.
Arensman B, Barrett J, van Bodegom A, et al. (2013) MFS II Joint Evaluation of International Lobbying and Advocacy. Baseline report. The Hague: NWO.
Arensman B, Barrett J, van Bodegom A, et al. (2015) MFS II Joint Evaluation of International Lobbying and Advocacy. The Hague: NWO.
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): 1310–26.
Armytage L (2011) Evaluating aid: An adolescent domain of practice. Evaluation 17(3): 261–76.
Coffman J (2011) Revolutionary lessons about interim outcomes. Advocacy Evaluation Update 11 (summer): 1–2.
Coffman J and Reed E (2009) Unique Methods in Advocacy Evaluation. Los Angeles, CA: The California Endowment.
Department of Social Affairs (2013) Interview with African Union Official. Addis Ababa, March.
Devlin-Foltz D, Fagen MC, Reed E, et al. (2012) Advocacy evaluation: Challenges and emerging trends. Health Promotion Practice 13(5): 581–6.
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.
Guba EG and Lincoln Y (1989) Fourth Generation Evaluation. Berkeley, CA: SAGE.
Gulrajani N (2011) Transcending the great foreign aid debate: Managerialism, radicalism and the search for aid effectiveness. Third World Quarterly 32(2): 199–216.
Hayman R, King S, Kontinen T, et al. (2016) Negotiating Knowledge: Evidence and Experience in Development NGOs. Rugby: Practical Action Publishing.
IOB Policy and Operations Department (2009) Evaluation Policy and Guidelines for Evaluations. The Hague: Netherlands Ministry of Foreign Affairs.
Keck M and Sikkink K (1998) Activists Beyond Borders: Advocacy Networks in International Politics. Ithaca, NY and London: Cornell University Press.
Lam TCM (2002) Theory-based evaluation and objective-based evaluation: An integration of the two approaches. Presented at the European Evaluation Society conference, Seville, Spain, 12 October.
Liket KC, Rey-Garcia M and Maas KEH (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): 171–88.
McDonald D (2008) Revisiting a theory of negotiation: The utility of Markievicz (2005) proposed six principles. Evaluation and Program Planning 31: 259–65.
Mansoor K (2003) Realist evaluation for practice. British Journal of Social Work 33: 803–18.
Marchal B, van Belle S and van Olmen 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 GE (1995) Ethnography in/of the world system: The emergence of multi-sited ethnography. *Annual Review of Anthropology* 24(1): 95–117.

Markievicz A (2005) ‘A balancing act’: Resolving multiple stakeholder interests in program evaluation. *Evaluation Journal of Australasia* 5(1&2): 13–21.

Organization for Economic Co-operation and Development (OECD) (1991) *Criteria for evaluating development assistance*. Available at: http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm (accessed 25 April 2016).

Patton MQ (2011) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: 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.

Putnam LL (2004) Transformations and critical moments in negotiations. *Negotiation Journal* 20(2): 275–95.

Riddell RC (2014) Does foreign aid really work? (keynote address). Presented 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): 413–36.

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.

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

Sharkey S and Sharples A (2008) From the beginning: Negotiation in community evaluation. *Evaluation* 14(3): 363–80.

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.

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

**Bodille Arensman** is a PhD candidate in the Strategic Communication and Humanitarian Aid and Reconstruction Chair Groups at Wageningen University & Research. Her research explores advocacy processes, evaluation and effectiveness, as well as the unpacking of the politics and challenges of advocacy. She was involved in the advocacy evaluation discussed in this article.

**Margit van Wessel** is an assistant professor in the Strategic Communication Chair Group at Wageningen University & Research. She explores the dynamics of communication and governance. Recent projects address citizens’ understandings of politics, advocacy for development and advocacy evaluation. She was the project leader and an evaluator for the evaluation discussed in this article.