Swiss Federal Institute of Technology Zürich

Department of Humanities, Social and Political Science

NADEL Center for Development and Cooperation

MAS in Development and Cooperation 2022-2024

Autumn Semester 2022

What are "emergent" approaches to measure the impact of complex and transformational organisational system-changes targeted at 'localising' Red Cross' International Cooperation work?

Concept Note

By:

Christian Fankhauser, Kingdom Karuwo and Mirjam Grünholz

In collaboration with:

Swiss Red Cross (SRC)

08.12.2022

Content

Summary

Localisation is a buzzword that has been circulating in the sectors of international humanitarian aid and development cooperation for decades. The discourse over the sectors' future has sparked in light of recent movements and events like the Black Lives Matter movement and the Covid-19 pandemic. Efforts and agreements, such as the Grand Bargain of 2015, have been joined by a multitude of proponents in the sector(s) that spotlight the prevailing power imbalances and challenge existing hierarchies between Global South and North.

The underlying challenge in this discourse is the diverging understanding of what localisation and locally-led development mean and how they should be implemented and achieved in practice. The most superficial form limits itself to decentralising organisations and employing local staff, while localisation in its genuine form questions the existing power structures and goes hand in hand with decolonisation.

While the structure of the Red Cross and Red Crescent Movement with its 192 National Societies is per se very decentralised, there exist fundamental differences among the National Societies concerning their capacities and organisational structure. To be fit for purpose and to account for the varying and changing environments within the Red Cross and Red Crescent Movement, an emphasis is put on National Society Development (NSD) to enhance mutual learning among sister organisations. Despite these aspirations, existing monitoring and evaluation measures fail to explain how and under what conditions NSD is facilitated sustainably, effectively and efficiently.

For these reasons, the concept note at hand wants to elaborate the following question: What are "emergent" approaches to measure the impact of complex and transformational organisational system-changes targeted at 'localising' Red Cross' International Cooperation work? The proposed solutions presented in the concept note at hand are twofold:

1) We elaborate why the existing **monitoring** framework used by the Swiss Red Cross to assess their National Society Development support needs an additional thematic priority to address localisation and the underlying power (im)balances between sister organisations. For this, we introduce three different tools, the Localisation Performance Measurement Framework (LPMF), the Local Engagement Assessment Framework (LEAF), and the Power Awareness Tool.

2) Moreover, concerning the unclear preconditioning factors for sustainably developing National Societies, we suggest complementing the continued monitoring with a profound **evaluation** through a Qualitative Comparative Analysis (QCA) that enables the identification of supporting and required factors for the investigated process through a rigorous qualitative measurement methodology.

Introduction

Within the International Red Cross and Red Crescent Movement (hereafter referred to as the Movement), the 192 member National Red Cross and Red Crescent Societies form the basic units and play the key role in delivering long-term services and development programmes on the one hand and disaster response and early recovery programmes on the other. Based on the Geneva Conventions[1], each signatory state agrees to the cooperation with the components of the Movement and to the aspiration to establish and develop a National Society (NS) on its territory, while providing the legal framework that assures this National Society's autonomy and defines its country-specific mandates (ICRC, 2006). Eventually, every National Society – through their branches – aims to be present in even the remotest communities performing local action.

All National Societies have developed more or less independently towards the shape they currently have, and thus, there is quite an organisational diversity in shapes, forms and core capacities. To strengthen the different National Societies – especially the ones existing in challenging environments – and to learn from each other, the *National Society Development (NSD) programme* of the International Federation of Red Cross and Red Crescent Societies (IFRC) was created (IFRC, 2013). Only through continued National Society development and constant adaptation of organisational structures can sustainability be achieved and the full potential of the Movement tapped.

Due to its great resource availability, the Swiss Red Cross (SRC) is in a position to contribute to the Movement wide NSD efforts and has been partnering with more than 30 National Societies in different projects on an annual basis and therewithin further collaborates with 20 sister Societies in the process of NSD support (SRC, 2022).

However, within the Movement as well as within SRC, there are profound and substantive doubts whether such NSD support – as it has been done up to now – is capable to create conditions that enable effective, efficient and sustainable organisational development, and whether the invested resources significantly contribute to improving the organisational capacities of those sister Societies. In this regard, the Swiss Red Cross strategy (SRC, 2020) explicitly puts emphasis in its 2030 strategy document on "working evidence-based", while "regularly reviewing the impact and necessity" of its work.

To date, various tools and assessment procedures have been used to try to address the measurement of the aspired organisational system changes. Nonetheless, up to now, no satisfactory approach has been developed and implemented to measure such complex, multidimensional, transformational, and non-linear changes. Well-established tools, such as the log frame, that follow a strictly linear logic in which predefined outcomes and outputs are measured against a set of indicators eliminate too much complexity and fail to provide accurate evidence of the effects of these complex processes. Hence, rethinking the assessment of complex organisational system change is necessary.

---

[1] The Geneva Conventions are a set of four treaties and three additional protocols, that constitute the legal framework for international standards for humanitarian treatment in war that have been ratified by 196 countries.

Furthermore, following the narrative of the Grand Bargain[2], of which the IFRC is one of 60 signatories, large humanitarian aid organisations and donors have acknowledged that there is a big need to shift their focus towards stronger and more effective local civil societies – or national societies in the case of the Red Cross and Red Crescent Movement – that shape their own priorities and strengthen their decision-making power. In a broader context, the core of this debate addresses the colonial past of development and tries to overcome it through localising funding and action and, ultimately, aid. Thus, power, power distribution and power imbalances must be reflected when assessing National Society Development and its measurement.

For these reasons, the concept note at hand wants to elaborate the question: What are "emergent" approaches to measure the impact of complex and transformational organisational system-changes targeted at 'localising' Red Cross' International Cooperation work?

The aim of the concept note is thereby twofold:

(1) It elucidates why – within the existing **monitoring** framework – additional indicators and measurement tools need to be implemented that address localisation through the lens of decolonisation. It outlines how this can be achieved in order that a better understanding of National Society Development (and its support) is enabled, and through the improved measurement and monitoring, the process towards strengthened National Societies can be supported more aptly and purposefully.

(2) Besides, this concept note wants to address the lack of appropriate measurement by promoting a robust qualitative methodological **evaluation** based on a Qualitative Comparative Analysis (QCA) that allows detection and assessment of what contributions of the SRC to National Society Development have been successful or not.

Problem Analysis

The call for localisation in humanitarian aid and development cooperation

The debate and discourse over the future of the humanitarian aid and development cooperation sectors has been sparked (again) over the last few years in view of the Covid-19 pandemic, the resurgence of the Black Lives Matter movement, as well as the increasingly felt effects of climate change (Baguios et al., 2021). Fundamental power imbalances between the Global South and North, like obstructive funding mechanisms and unequal decision-making systems, are widely criticised and more vocally spoken about.

The importance of *localisation* and *locally-led development* is recognised to be at the core of the debate and has found its way into agreements like the Grand Bargain. With such agreements, prominent development actors hope to create guidance for organisations wishing to partner at various stages and levels with local actors

---

[2] The Grand Bargain is an agreement between some of the largest donors and humanitarian aid organisations with the goal of reforming humanitarian aid delivery (national governments as well as NGOs are signatories). It was signed in May 2016 at the World Humanitarian Summit and targets were supposed to be met by January 2020 (IASC, 2022).

(Derzsi-Hovath et al., 2017), to build capacity in local contexts (Eade, 2010; Ika & Donnelly, 2017) and to channel resources for development in a meaningful and impactful manner (Muchadenyika, 2016).

The Grand Bargain agreement is rooted not only in the debate over localisation, but is also a consequence of severe shortcomings in past humanitarian aid missions and the resulting need of improvement in aid delivery. Even though the Grand Bargain has its origins in humanitarian aid, the debate goes way beyond it and is equally fundamental for development cooperation and international cooperation in general.

However, despite some willingness of adopting locally led development practises – as exemplified by various anecdotal success stories (e.g. iHub movement in Africa, Petajakarta project in Indonesia, UNICEPF's Innovation Lab (cf. Wall & Hedlund, 2016)) – the 2020 Annual Independent Report for the Grand Bargain has acknowledged that the practical implementation of the localisation agenda is yet to be meaningfully achieved (Peace Direct, 2021) and the commitments and promises of the Grand Bargain, especially concerning the topics of localisation (and the ominous 25% of funding going to "local or national organizations"), are far from being met (Baguios et al., 2021)

What is localisation – and why is it so difficult to implement?

The term localisation is not without controversy in both its conceptual and – as outlined in the previous chapter – in its practical meaning. In the discussion that led up to the Grand Bargain, the term 'localisation' was often used – and even more contested. 'Localisation' has become an umbrella term to describe any activity involving local actors, such as, but not limited to, national governments, local and non-governmental organizations (NGOs), civil society and communities of the Global South (Wall & Hedlund, 2016).

Localisation can be viewed as the journey or path, while locally-led practice or development constitutes the goal thereof. Advancing this thought, if locally-led-practice is the ideal to strive for, then localisation incorporates all measures and steps that are necessary for this process to be successful. By definition both concepts are strongly interlinked and any transformation process interacts with the existing practice, yet, the distinction and conceptual separation can help to understand the underlying power dynamics (Baguios et al., 2021). Following this viewpoint, the risk of confounding the path with the end goal is as dangerous as the fallacy of believing that the path of localisation necessarily will lead to the destination of the sector 'being locally led' (Baguios, 2021).

The confusion between localisation (as being the path) and locally-led development (being the goal) is not the only problem as also the interpretation of the eventual goal of development can vary strongly. Figure 1 exemplifies the view of an actor from the Global North displaying the spectrum of different intensities of 'locally-led' ranging from information provision to supporting initiatives by local actors. In this context, locally-led development (LLD) can be described as the process in which local actors "set their own agendas, develop solutions, and bring the capacity, leadership, and resources to make those solutions a reality" (USAID, 2021). However, a closer look reveals the shortcomings of this stance. In the past decades, international actors have often remained at the centre of the debate hiding behind terms such as localisation or locally-led

development to conceal the prevailing power imbalance (Barbelet et al., 2021). Partnering with local actors was of advantage to use them as sub-contractors and outsource fiduciary and security risks (Brugger et al., 2022). Nourished by a lack of trust and general risk averseness of donors, the perception has remained that collaboration on eye-level would increase the risk of passing a project cycle unsuccessfully, and if so, losing reputation and putting future funding at risk (Barbelet et al., 2021). Moreover, collaboration was constrained by the perception that local actors could not operate at scale, are not able to attract sufficient funding and meet due diligence requirements (Venton, 2021).
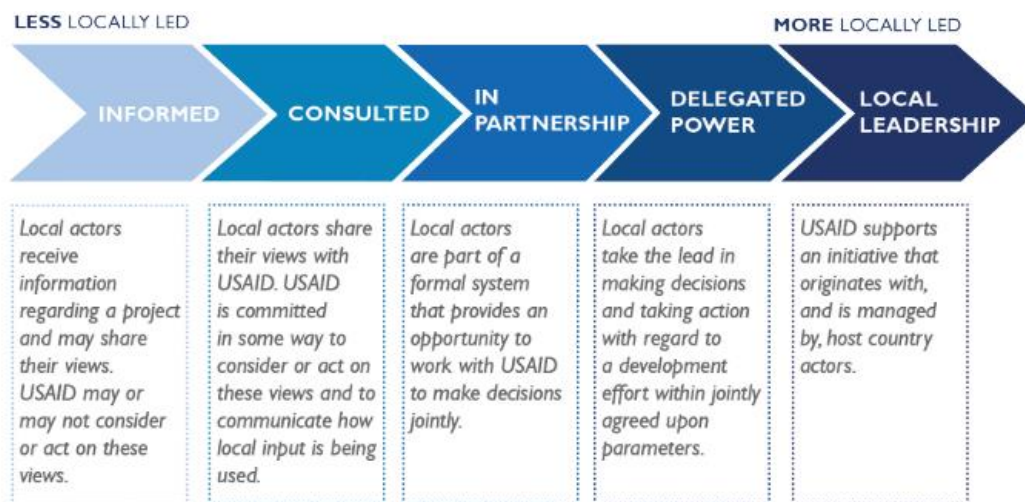

Figure 1: Global North views on the forms of locally-led development (USAID, 2021)

Trying to avoid the mistakes of the past and to not cling to the predominant top-down approaches, it is crucial to shift the spotlight onto the viewpoints of actors of the Global South and, in this line, focus on the existing power imbalances between Global North and South. The fundamental problem of power (im)balances and the undeniable reciprocal relationship to localisation is emphasised in the following quote:

> *"In its most basic form, power is an obstacle to localisation because power remains in the hands of international actors to decide who has capacity or not, what capacity counts, who gets funding or not, what types of partnerships prevail, and who gets access to coordination structures and strategic decision-making forums."*

Barbelet et al. (2021)

Only a combination of the perspectives of the Global South and the Global North enhanced with a deep focus on power can help us understand the association between imagined development partnership and the specific implementations thereof, and ultimately what locally-led development should aim at.

Local actors are often nudged into accepting development as it is, or risk losing the very much needed support without which they would not be able to exist and sustain themselves, due to limited power and lacking financial resources, therefore rendering them unable to control the development discourse (Venton, 2021). This dominance and wielding power of actors predominantly from the Global North often leads to

sentiments of frustration and reinforce the colonial allures that remain subject of much criticism by local actors (Barbelet et al., 2021; Barnett et al., 2005). To illustrate, let us consider the quote below:

> *"I, my colleagues and many other local actors alike are not stupid. We actually know stuff.*
> *And guess what? We are capable of knowing more! We recognise that we are not perfect.*
> *Many of us were educated in a country lacking responsible leadership and without much access*
> *to information and opportunities. But I also know my community and I am working so hard*
> *to be a part of driving the change that we want to see in our society"*

Usen (2019)

This quote adds a significant number of issues, chief among which is a lack of recognition of knowledge, competence, and admission of the challenge of being a development practitioner from the Global South. Development assistance should therefore delve deeper than just setting up local offices and hiring local staff; it should include the vision of what development is in a given polity (Ang, 2018). There is reasonable evidence to buttress the idea that some level of commitment on the side of donor entities to commit to sharing, shifting and/or balancing power could be a worthwhile pursuit and goal. Thus, without a genuine shift of power, localisation may only localise (or decentralise) the sector without incorporating the ideals of being truly locally-led (Baguios, 2021).

In this discourse, compelling questions refining the concept of localisation – such as who defines development (Schuller, 2012), what is local (Appadurai, 1988), and what is community (Ledwith, 2020) – have all emerged to address the normative relationship between development actors. The ultimate idea is that the regions with most resources must also allow agency and self-determination of regions requiring development assistance and that such assistance must take into account practice, context, culture and value of those 'local'.

Furthermore, critics of its practical implications of localisation have expressed the concern that to accompany the various dictates of what localisation means would require a deeper commitment to dealing with the remnants of the helper syndrome resonant – both historically and economically – as well as with the very recent colonial epoch (Rodney, 2018). Calls for decolonising aid and overcoming power imbalances hence echo throughout the sector with hope to undermine the perpetuation of colonial strokes which dominate the relationship between donor entities and their local so called 'partners'. The latter increasingly demand some aptitude to power to determine their development destiny. They strongly reflect on the uneasy historical relationship between the today Global North and Global South regions which are rooted in unequal decision-making, racist and colonial structures including a perpetual system of dominance of the North over the South subsequently argued to be leading to an anti-developmental dependency (Moyo, 2010).

Thus, despite this ubiquity in the sector, localisation remains elusive and lacks a modality on which to base practice. Consequently, the Swiss Red Cross must face the question of how it can, as a humanitarian organisation, deal with questions of decolonisation more broadly and localisation specifically, how and by what means development practise can be decolonised and whether the efforts for more locally-led

development can achieve this goal without risking funding. All this should lead to actions to improve lives in the Global South, and ultimately enhance development cooperation partnerships while facing the challenges arising in implementing and assessing localisation.

Strategy 2030 of the International Red Cross and Red Crescent Movement

The structure and organisation of the Red Cross and Red Crescent Movement inherently are very decentralised, and its work – through the principle of volunteering – relies strongly on the engagement of local actors. Yet, it has been observed that, in reality, distinct power imbalances exist that need to be addressed. To that effect, the *Strategy 2030* of the International Federation of Red Cross and Red Crescent Societies (IFRC) repeatedly "proposes an urgent shift of leadership and decision-making to the most local level – placing local communities at the very centre of change" (IFRC, 2018). The strategy further refers to the Grand Bargain and its commitments while identifying seven transformations to be essential for the Movement. For example, the first aspired transformation to ensure strong and effective National Societies states that a "much greater focus [will be put] on the development of branches and National Societies, ensuring services are led and developed by local actors, and that National Societies have a stronger role in setting their own priorities and a stronger voice in any decisions being made about operations in their own territory" (IFRC, 2018). Later on in the same document it is explained that "our first accountability is to the communities we serve" and "we must also continue to support the independence of local actors to enhance their decision-making and agency" (IFRC, 2018).

All further guiding documents of the IFRC that focus on National Society Development and the support thereof, like the *NSD Framework*, *Compact* and *Policy*, reiterate this notion of a necessary transformation towards locally-led projects and the decision power being at the lowest level possible (IFRC, 2013, 2019, 2022). In addition, reoccurring themes like the ownership of the development of National Societies and their inevitable uniqueness, as well as the accountability of a National Society to "the people and communities it engages with and with whom it develops services and programmes" stress the necessity of a profound localisation process (IFRC, 2022).

The challenge of measuring localisation

The United Nations 2030 Agenda for Sustainable Development recognizes that capacity-building is a substantial component of the means of implementation for the Sustainable Development Goals (SDGs). Particularly SDG 17 concentrates on *localisation* in the form of resource mobilisation, knowledge sharing, enhancing capacities and encouraging effective partnerships. National Society Development, capacity enhancement and organisational development are thus not goals in themselves, but are an effort aiming to achieve and maintain accountable and sustainable organisations that deliver relevant services to address needs, reduce vulnerabilities and build resilience in changing environments.

As has been shown, development practices of the SRC aspire to be in line with SDG 17 and aim at becoming locally-led. Nonetheless, in reality major challenges persist in the actual implementation. Despite the decentralised structure of the Movement and SRC's mandate to support other NSDs in their development and institutional capacity building, it remains unclear how effective, efficient and sustainable the activities conducted by SRC in this regard are. Moreover, in line with the Grand Bargain, there is an increasing demand for compelling evidence to foster and enrich argumentation and negotiation for sustained investments into local actors allowing them to go through the desired change processes and build the capacities they need.

The IFRC as well as SRC recognise that NSD and the underlying organisational system-changes are highly complex processes and generally do not follow a strictly linear pathway. Therefore, established methods for impact measurement in which predefined outcomes and outputs are measured against set indicators in a Log Frame structure are falling short in providing actual evidence of the effects (intended or unintended) of such processes. In consequence, the question arises how the SRC can assess to what extent the NSD activities undertaken by them strengthen their sister organisations and support them in maintaining their capacity and implement strategies that minimize the negative impacts of current social, economic and environmental crises and emerging challenges.

Walking the talk – call to action for the Swiss Red Cross

It is our view, that in light of the debate about effective implementation and assessment of National Society Development – and ultimately locally-led development and decolonisation – the Swiss Red Cross could add various facets responding to such criticism.

There is on the one hand a need for **improved monitoring tools regarding the degree of localisation** for the further progress in and improvement of the decolonisation process through deeper understanding and the creation of data – within the Movement as well as externally in the sector. On the other hand, additional **evaluative measurement approaches that account for the complexity of National Society Development processes** and that can give insights into the effectiveness of NSD support are required.

To account for this, firstly, we recommend that the Swiss Red Cross integrates an additional evaluative lens in their **monitoring** process which assesses the extent of localisation in their partnerships. It is herein necessary, we argue, to not limit localisation to one sagacity and narrative, because the discourse is huge and there stands to be gains in adopting it broadly, since any progress in that direction must be viewed as worthwhile and trend setting. Through the incorporation of measurement tools that pinpoint the degree of localisation value can be created for the Red Cross and Red Crescent Movement.

Secondly, we recommend a profound assessment method for **evaluation** to identify preconditions that are critical for National Societies to develop, and to assess to what degree the Swiss Red Cross is able to considerably enhance and contribute to these processes. Such an evaluation can help in the learning process of what it is that works best in different areas and what does not, while possibly providing scalable results and findings not only for within the Movement, but also externally.

Proposed Solution

Overview

To answer the question what "emergent" approaches exist to measure the impact of complex and transformational organisational system-changes targeted at 'localising' Red Cross' International Cooperation work, we propose two solutions, one addressing the monitoring and one addressing the evaluation of SRC's interventions in National Society Development.

Regarding **monitoring**, we propose to expand the existent thematic priorities in the "NSD Indicator Toolbox" of the Swiss Red Cross by adding 'Localisation' as a thematic priority. More specifically, the Power Awareness Tool and other related tools are proposed as emergent evaluative monitoring tools. These tools can be used as a baseline to conduct yearly monitoring of the mutual relationship between the SRC headquarters and National Societies that SRC is mandated to support in their development.

Regarding **evaluation**, we suggest complementing the continued monitoring by one large evaluation that will complement with its findings to the knowledge generated throughout the years. There are many qualitative evaluation tools that can be taken into account, but given the challenge and the global dimension thereof, we recommend conducting a Qualitative Comparative Analysis (QCA). This decision is founded on our understanding of the challenge and supported by an assessment of different evaluation approaches, using the "Choosing Appropriate Evaluation Methods Tool" provided by BOND (BOND, 2016). The result of the conducted assessment is attached at the end in the Appendix.

In the following sections, both solutions will be described in more detail.

Localisation measurement tools

Before delving into the fathom of QCA, a condensed summary of tools available for assessing how far an organisation is on the path of localisation, is given. The dedication for locally driven development, among other things, has led to a considerable number of recommendations on how it can be implemented, measured, evaluated, and even monitored.

From the plethora of approaches available, we will discuss three novel tools:

a) Localisation Performance Measurement Framework (LPMF)
b) Local Engagement Assessment Framework (LEAF)
c) Power Awareness Tool

a) Localisation Performance Measurement Framework (LPMF)

According to the Network for Empowered Aid Response (NEAR), the purpose of the Localisation Performance Measurement Framework (LPMF) is to demonstrate progress in meeting localisation obligations (Featherstone, 2019). The focus is on local and national stakeholders, but it is also expected to be

relevant to international non-governmental organisations, UN agencies, donors, and research and academic institutions evaluating localisation. The development of the LPMF has been guided by the goal of providing a clear, practical tool that can help strengthen the evidence base to locate and promote a common understanding of progress made and also identify weaknesses.

The LPMF seeks to assess localisation within the following six conceptual components: 'partnerships', 'funding', 'capacity', 'coordination and complementarity', 'policy, influence and visibility' and 'participation'. Each component has an over-arching desired change and comprises multiple groups of key performance indicators which need to be assessed through different means of verification (qualitative or quantitative measures), wherefore the Framework suggests appropriate measurement strategies. At the end of the assessment process an overview table summarizes the localisation progress and guides the critical reflection and development of an action plan.

b) Local Engagement Assessment Framework (LEAF)

The Local Engagement Assessment Framework (LEAF) was designed by Oxfam America, Save the Children and the Overseas Development Institute as a resource for the planning and evaluation of the level of involvement of local stakeholders in development projects. LEAF was constructed as a qualitative planning tool to help development professionals design, monitor and evaluate projects to integrate solid tenure principles. LEAF's goal is to encourage the highest level of local ownership that is appropriate, based on the project and context, at each stage and in each part of a single project (Oxfam, 2017).

LEAF facilitates an assessment of ownership by graphically representing which local stakeholders under the host government, civil society and the private sector were involved in a project, while displaying if such involvement allowed a meaningfully co-shaping of the project and during which parts of the project cycle the engagement took place. The application of the LEAF follows a step-wise procedure, where first the stakeholders are identified, before the degree of local stakeholders' ownership of and engagement in the topics of 'priority setting', 'implementation', 'resources' and 'sustainability' is assessed. The engagement levels (in each phase of a project) are assessed on basis of a four-stage scale ranging from 'informing', 'consultation' to 'partnership' and 'delegated power'.

Since it is a monitoring and evaluation tool, LEAF can capture which stakeholders were involved, but not which ones should have been. To that effect it is well-equipped to compare projects and map an organisation's progress in the journey of localisation.

c) The Power Awareness Tool

The Power Awareness Tool is intended for organisational reflection to determine the extent to which an organisation is aware of its power; with the primary objective being to use power to empower and not to perpetuate dominance.

According to Partos (2022) is has three simple steps in its utilisation, namely:

1. Identification of important decision-making topics
2. Scoring the level of participation of each partner in a decision-making process
3. Reflection on steps 1 and 2.

It is important that the facilitator creates the right atmosphere and preconditions before the steps begin. Critical feedback should not be feared by participants. The only way they will do so is, if they feel that the intentions behind organizing a power awareness session are genuine, and that the intention is to achieve a more balanced power relationship within the partnership (Partos, 2022). The assumption is that if the partners have an understanding of how power works in the partnership, they will be able to work better to shift power in accordance with shared principles. Power imbalances can lead to misunderstandings, feelings of unfairness, friction, and loss of trust, that ultimately can lead to projects not being efficiently conducted or even aborted. Therefore the tool emerges as an important and novel one to interrogate power and ultimately decolonise an organisation (Elbers, 2012).

Power Awareness Tool in practise

As broadly highlighted above, the Power Awareness Tool is endowered with monitoring prowess which renders it useful for organisations – such as the Swiss Red Cross – who wish to strengthen their partnerships by unfolding and overcoming underlying power relations. More specifically, the tool can be used to assess power structures between the SRC and a partnering National Society, but also within SRC or said National Society separately. Let us consider the steps more deeply, with a focus on the evaluative impetus.

1. **Identification of important decision-making topics**

The first step in applying the tool requires teams – or stakeholders – to identify key decision topics that are emerging in their partnerships or that may emerge in new partnerships. They should be listed in column 1 of the power analysis framework as demonstrated in Figure 2 below (Elbers et al., 2018; Kilby, 2006).

This step primarily recognises that issues have varying priorities to different actors and accordingly seeks to provide a platform upon which discussions of what is imperative can be tabled. Such decisions may include a wide range of subjects like budget allocations, recruitment issues and more importantly, choosing one program over a competing one. Put simply, the first step is concerned with making sure that the diverging views on what is important for different partners can be represented. This step primarily challenges hierarchical top-down approaches to decision-making by prioritising all the views of everyone involved in tentative implementation planning. Elbers (2012) reclaims that it is important to list every decision-making topic even if it may be considered important only by one partner, a philosophy which resonates strongly with the United Nations Sustainable Development Group (UNSDG) resolution of "leaving no one behind" (UNSDG, 2022). The exact modalities of the smaller but important stages in this step are self-explaining and demonstrated quite illustratively in the same figure below.

Power analysis framework

Key decision-making topics

Level of participation
4 Partner decides
3 Partner co-decides
2 Partner is consulted before a decision is made
1 Partner is informed about decision-making
0 Partner is not involved in decision-making

	Partner A	Partner B	Partner C	Partner D	Partner E	Partner F	Partner G
Scoping and building							
1							
2							
3							
Etc.							
Managing and maintaining							
1							
2							
3							
Etc							
Revisiting and revising							
1							
2							
3							
Etc							
Sustaining outcomes							
1							
2							
3							
Etc.							
Total	Σ	Σ	Σ	Σ	Σ	Σ	Σ

Figure 2: Power analysis framework – template table (Partos, 2022)

2. Scoring the level of participation in decision-making

Even if step 1 is done, there is no way of assessing if it was done in a manner that satisfied the partners involved, hence step 2 bears crucial importance. In step 2, the idea is for stakeholders to score each partner on their relative power to influence key decisions tabled in step 1, and this is done by using the ladder of participation in decision-making outlined in Figure 2 above. By tallying the scores, the degree of participation by each partner on a given decision is determined. A scale can be decided upon, for example (0) meaning not involved and (5) meaning highly involved. After the scores are added for all different decisions, a greater score may indicate that the partner has a lot of power, while a lower score demonstrates the contrary. It is, however, important to note that some partners may have higher scores on certain decision-making topics, and low scores in relation to topics of less importance. While this reveals that topics have different magnitudes of importance and that responsibilities differ across partners, it is nevertheless useful to think about why that is the case, and this then initiates and necessitates the third and final step (Fowler, 2000).

3. Reflection

Step 3's primary objective is to reflect on the power (im)balance in a partnership after it has been made visible through scoring the level of participation in relation to key decision-making topics. To buttress the succinctness of this stage, Fowler (2000) recommends the following guiding questions and argues that they can act as an orientation for reflection:

With regards to column 1:

> Are these decision-making topics relevant to your partnership, or are there additional topics which must be included in the analysis?

With regards to the other columns:

> Are all partners content with their level of participation or should their level of participation be up- or downgraded and why?
>
> If a higher degree of influence of a partner on a particular topic is deemed necessary or desirable, would the partner actually be more influential, or is the partner unable to exact more influence, and what can be done to alleviate this problem?

Potential limitations and shortcomings of the Power Awareness Tool

Like any methodological apparatus, perfection is a mirage; there are aspects which are beyond the whelm of this Power Awareness Tool and these are also important to point out.

While the tool can be pivotal and instrumental to check power (im)balance of existing partnerships, the same cannot be said confidently for new partnerships. In existing partnerships, it is useful to check – on a regular basis – whether or not partners are still content with the way decisions are made. New partnerships, however, require greater sophistication primarily because of pre-existing power dimensions which could be inherent to international cooperation in general (Elbers et al., 2018). Yet, even with this potential limitation, it qualifies as innovation to deploy the Power Awareness Tool in various stages of the partnering cycle to undermine the prevalence of contextual challenges reeling development cooperation partnerships.

Qualitative comparative analysis (QCA)

Complementary to introducing localisation measurement tools into the existing monitoring framework of the NSD support of SRC, we suggest the conduction of a one-time evaluation using a Qualitative Comparative Analysis (QCA) that will complement to the knowledge generated throughout the years.

Qualitative comparative analysis is a research method that has increasingly found its way into monitoring and evaluating of development projects and programs. It is particularly beneficial for organisations that run multiple projects in different places and want to assess their effectiveness. In addition, the method is suitable for analysing multiple cases in complex situations and can help explaining why change happens in some cases but not in others (Simister & Scholz, 2017).

QCA is a rigorous method that relies on a relatively small case base of usually 10 - 50 cases. It assesses *whether* an intervention made a difference and *through which patterns* this difference was achieved. To answer this, QCA draws on qualitative (case-oriented) and quantitative (variable-oriented) approaches. It requires an in-depth and thorough understanding of each case, which can be obtained through qualitative research methods such as interviews, observations, or literature reviews (Schatz & Welle, 2016). The factors identified in the qualitative part are then translated into a quantitative numerical format which allows a systemic analysis of patterns across the data (Schatz & Welle, 2016). QCA results in a list of conditions that should be fulfilled if a certain change is desired to happen. If these conditions are fulfilled in another context, e.g., if a young National Society is developed, the likelihood that the intended actions will be successful increases once the identified conditions are present (or absent). As such, QCA is capable of uncovering underlying cross-case patterns considering the diversity of cases and the heterogeneity of contexts by comparing cases as configurations (Ragin, 1984).

QCA has the advantage, that it is a rigorous and transparent method that can easily be replicated or tested by others (Simister & Scholz, 2017). It goes beyond many case-based approaches, which are often criticised for lacking generalisability and replicability, and that are considered to "only" provide anecdotal evidence. Instead, it is based on a pre-existing, or as part of the QCA developed, clear and detailed theory of change of projects or programs, and identifies what conditions are necessary or sufficient for an intervention to achieve a particular outcome. It enables a systematic comparison of cases based on identified key factors which are responsible for the success or the failure of an intervention. Thus, it bears the potential to be strong regarding its internal and its external validity (Baptist & Befani, 2015). Furthermore, QCA acknowledges that change may not be linear but is often induced through multiple combinations of factors rather than an individual factor, that different combinations of those can result in similar changes and that one factor may have different impacts on an outcome, depending on the context and its combination with other factors (Rihoux et al., 2011). Ultimately, the method can take into account conditions that change over time and across actors or cases (e.g. policy arenas), as well as conditions that are stable over time (e.g. institutions, context, system) (Korhonen-Kurki et al., 2014).

In Monitoring and Evaluation, the methodology can be used for understanding *how* and *why* specific results are achieved or not achieved. To do so, a set of essential requirements need to be fulfilled. Firstly, QCA needs a clear theory of change that includes detailed key factors that are "felt" to be required for a project intervention to achieve a particular outcome or impact. Secondly, based on the theory of change, a set of key factors as well as their presence or absence in each case needs to be assessed. These factors need to be compared carefully and qualitatively benchmarked based on some pre-defined scoring criteria (Baptist & Befani, 2015). This assessment can be binary (0/1), resulting in a *crisp-set QCA*, or more fined-grained with scores ranging from 0 to 1 (e.g. 0, 0.33, 0.66, or 1), resulting in a *fuzzy-set QCA* (Simister & Scholz, 2017). Thirdly, clear, distinct cases must be present. They must show different levels of "success", i.e., some cases must show positive outcomes, while others show negative ones, so that factors leading to "success" or "failure" can be identified and compared[3]. Finally, QCA requires some technical capacity and understanding of Boolean logic, set theory and QCA software (Baptist & Befani, 2015). Different software is available to conduct QCAs, such as the free computer software fsQCA (Ragins, 2022) , TOSMANA or several R-based packages such as QCA, QCA3 and QCApros (Compasss, 2022). Once these requirements are fulfilled, this complex QCA approach accounts well for development interventions that usually work differently in different contexts or geographical areas, and for which there is no "one way fits it all" approach that guarantees success (Baptist & Befani, 2015). QCA can therefore ultimately be used to learn from one's own work, to understand past interventions better, inform future planning, and display accountability to donors (Simister & Scholz, 2017).

There are, however, a few weaknesses that need to be recognised. QCA needs a minimum number of cases to work effectively and cannot cope with lacking data. In case of missing data, factors or entire cases need to be excluded. Scoring single factors requires a solid understanding of each case and complex judgment. Besides, QCAs are quite time- and resource-intensive due to the iterative evaluation of the data, but nonetheless, QCA can be conducted without much prior knowledge, given that the team is willing to spend some time and resources on learning, trial and error (Simister & Scholz, 2017).

---

[3] Here it is important to mention that the level of "success" or "failure" is based on rigorous methodological case study comparison and serves the organisation to better understand what combination of factors – or what type of support – contributes to a successful outcome and what not. Therefore, we recommend the organisation to distance itself from any normative understanding of "good" or "bad", "successful", or "not successful". Instead, we suggest thinking more broadly about the advantages of such an impact evaluation and how it can contribute to the question of whether interventions make a difference and through which interventions differences are achieved.

Conclusion

The Swiss Red Cross is mandated to support the International Red Cross and Red Crescent Movement in National Society Development of its sister organisations. Hereby, these sister National Societies are constantly developing and adjusting to ever-changing environments and circumstances – and so is the SRC. However, to what extent the activities of the SRC contribute to these complex and transformational organisational system changes remains open. For this reason, the SRC has been looking for emergent approaches to assess these changes with the goal of localisation of the work of the Red Cross' international cooperation.

Focusing on localisation and locally-led development, we outlined the debate from a theoretical and normative perspective, combining arguments from the Global North and the Global South. We differentiated between localisation as the path and locally-led development as the goal and by doing so, we set the arguments in the broader context of power (im)balance and decolonisation. While localisation implies the transfer of power at different degrees from powerful international organisations and donors predominantly presiding in the Global North to actors of the Global South, there is still a widespread feeling of discomfort in handing over all responsibility and duties. Yet, exactly this would be necessary to build trust, strengthen local capacities, and transform the so-called "beneficiaries" into fully recognised, valued, and appreciated partners.

Nonetheless, the challenge remains of defining how one can measure the degree of localisation and power (im)balances of and within organisations. Tools like the conventional Log Frame follow a strictly linear logic which can be of value for implementing a broad range of development projects. However, they fail to monitor and evaluate transforming organisational system-changes in a highly complex environment – such as the ones where the National Societies of the Movement are embedded in.

For this reason, we proposed two solutions to face these challenges: Firstly, we recommended considering localisation as an additional priority area in the existing monitoring framework of National Society Development support of the SRC by introducing three different tools, i.e., the Localisation Performance Measurement Framework (LPMF), the Local Engagement Assessment Framework (LEAF), and the Power Awareness Tool that assess the degree of localisation and uncover existing power hierarchies. Secondly, we advanced a robust methodological evaluation method, a Qualitative Comparative Analysis (QCA), which allows the comparison of multiple cases in a complex environment and can help explaining why change happens in some cases but not in others.

References

Ang, Y. Y. (2018). Going Local 2.0: How to Reform Development Agencies to Make Localized Aid More Than Talk. *Stanford Social Innovation Review*, 2003–2005.

Appadurai, A. (1988). Introduction: Place and Voice in Anthropological Theory. *AmericanAnthropological Association*, *3*(1), 16–20.

Baguios, A. (2021). *Localisation Re-imagined: Localising the sector vs supporting local solutions*. Third World Quarterly; ALNAP Active Learning Network for Accountability and Performance in Humanitarian Action.

Baguios, A., King, M., Martins, A., & Pinnington, R. (2021). *Are we there yet? Localisation as the journey towards locally led practice: Models, approaches and challenges. October.*

Baptist, C., & Befani, B. (2015). Qualitative Comparative Analysis - A Rigorous Qualitative Method for Assessing Impact. *Coffey, June*, 1–7.

Barbelet, V., Davies, G., Flint, J., & Davey, E. (2021). Interrogating the evidence base on humanitarian localisation A literature study. In *HPG Literature Review* (Issue June).

Barnett, M., Duvall, R., Wendt, A., Crawford, N., Sikkink, K., Kinsella, H., Finnemore, M., Weldes, J., Pevehouse, J., Hurrell, A., Ruggie, J., Snidal, D., Keohane, R., Kupchan+, C., Burch, K., Diez, T., Donahue, T., Duvall, W., Gundogdu, A., … Slaughter, A.-M. (2005). Power in International Politics. *International Organization*, *59*, 39–75.

BOND. (2016, October 16). *Choosing Appropriate Evaluation Methods Tool.* https://www.bond.org.uk/resources/evaluation-methods-tool/

Brugger, A. F., Holliger, J., & Mason, S. J. A. (2022). *Triple Nexus in Fragile Contexts : Next Steps.*

Compasss. (2022). *Software | COMPASSS*. Comparative Methods for Systematic Cross-Case Analysis. https://compasss.org/software/

Derzsi-Hovath, A., Steets, J., & Ruppert, L. (2017, June 19). *Independent Grand Bargain Report*. Global Public Policy Institute; Routledge.

Eade, D. (2010). Capacity building: who builds whose capacity? *Development in Practice*, *17*(4–5), 630–639.

Elbers, W. (2012). The Partnership Paradox: Principles and Practice in North-South NGO Relations. In *Spiritus A Journal of Christian Spirituality*.

Elbers, W., Frobisher, E., Kamau, P., Kumi, E., Saharan, T., & Schulpen, L. (2018). *Aid chains and advocacy in the Global South : asset, nuisance or necessary evil? : literature review.*

Featherstone, A. (2019). Localisation performance measurement Framework. In *Network for Empowered Aid Response (NEAR)*.

Fowler, A. (2000). PARTNERSHIPS: NEGOTIATING RELATIONSHIPS - A Resource for Non-Governmental Development Organisations. *Occasional Papers Series Number 32*, 1–23.

IASC. (2022). *About the Grand Bargain*. https://interagencystandingcommittee.org/about-the-grand-bargain

ICRC. (2006). *Statutes of the International Red Cross and Red Crescent Movement.*

IFRC. (2013). *National Society Development Framework 2013.*

IFRC. (2018). *Strategy 2030: A Platform for Change.* https://future-rcrc.com/s2030/

IFRC. (2019). *National Society Development Compact 2019: A collective commitment to a global network of strong local actors.*

IFRC. (2022). *National Society Development: Policy* (Issue June).

Ika, L. A., & Donnelly, J. (2017). Success conditions for international development capacity building projects. *International Journal of Project Management*, *35*(1), 44–63.

Kilby, P. (2006). Accountability for Empowerment: Dilemmas Facing Non-Governmental Organizations. *World Development*, *34*(6), 951–963.

Korhonen-Kurki, K., Sehring, J., Brockhaus, M., & Di Gregorio, M. (2014). Enabling factors for establishing REDD+ in a context of weak governance. *Climate Policy*, *14*(2), 167–186.

Ledwith, M. (2020). *Community development : a critical and radical approach* (3rd ed.). Policy Press.

Moyo, D. (2010). *Dead Aid: Why aid is not working and how there is a better way for Africa*. Farrar, Straus and Giroux.

Muchadenyika, D. (2016). Multi-donor Trust Funds and Fragile States: Assessing the Aid Effectiveness of the Zimbabwe Multi-donor Trust Fund. *Journal of International Development*, *28*(8), 1337–1357.

Oxfam. (2017). *A practitioner's guide to integrating country into development projects*.

Partos. (2022). *The Power Awareness Tool*. https://www.partos.nl/publicatie/the-power-awareness-tool/

Peace Direct. (2021). *Time to Decolonise Aid Insights and lessons from a global consultation Time to Decolonise Aid About this report*.

Ragin, C. C. (1984). *What is Qualitative Comparative Analysis (QCA)?*

Ragins, C. (2022). *fsQCA Software*. https://www.socsci.uci.edu/~cragin/fsQCA/software.shtml

Rihoux, B., Rezsöhazy, I., & Bol, D. (2011). Qualitative Comparative Analysis (QCA) in Public Policy Analysis: an Extensive Review. *German Policy Studies*, *7*(3), 9–82.

Schatz, F., & Welle, K. (2016). Qualitative Comparative Analysis: A Valuable Approach to Add to the Evaluator's Toolbox? Lessons from Recent Applications. *CDI Practice Paper*, 1–8.

Schuller, M. (2012). *Killing with kindness: Haiti, international aid, and NGOs*. Rutgers University Press.

Simister, N., & Scholz, V. (2017). Qualitative comparative analysis (QCA). In *M&E Universe*.

SRC. (2020). *SRK-Strategie 2030*.

SRC. (2022). *National Society Development*.

UNSDG. (2022). *UNSDG | Leave No One Behind*. https://unsdg.un.org/2030-agenda/universal-values/leave-no-one-behind

USAID. (2021). *What is locally led development?*

Usen, L. E. (2019). Localisation: we are frustrated, not stupid! *FM Review*, 1–2. www.agendaforhumanity.org/initiatives/3861

Venton, C. C. (2021). *Direct support to local actors: considerations for donors* (Issue August).

Wall, I., & Hedlund, K. (2016). *Localisation and Locally-led Crisis Response: A Literature Review*.

Appendix

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

_____

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

> What are "emergent" approaches to measure the impact of complex and transformational organisational system- changes targeted at 'localising' Red Cross' International Cooperation work?

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
Fankhauser	Christian
Karuwo	Kingdom
Grünholz	Mirjam

With my signature I confirm that
- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

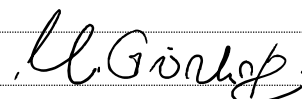I am aware that the work may be screened electronically for plagiarism.

Place, date

Zurich, 07.12.2022

Signature(s)

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.

STAGE 1: Which evaluation method(s) are suited to answering your key evaluation questions?

Do you want your evaluation to answer the following Key Evaluation Questions?	User Response	RCT (Randomised Contol Trial)	Difference-in-Difference	Statistical Matching	Outcome Mapping	Most Significant Change	Soft Systems Modelling	Causal Loop Diagram	Realist Evaluation	QCA (Qualitative Comparative Analysis)	Process Tracing/ Bayesian Updating	Contribution Analysis
1.1 Do you want to know "What was the additional/ net change caused by the intervention?" *or* "How much of the observed outcome(s) can be attributed to the intervention?"? *(Note: this is the core question of experimental and quasi-experimental impact evaluations)*	Yes	5	4	3	1	2	1	1	1	2	1	1
1.2 Do you want to know "What difference did the intervention make to different population groups, and under what circumstances?" (i.e. you are interested in effects for different groups and contexts, not just an "average" effect)	Yes	4	4	4	4	3	4	4	4	5	3	4
1.3 Do you want to know "How and why did the intervention make a difference, if any? or What was the process/ mechanism by which the intervention led to or contributed to outcomes?" *(Note: this is typically the main focus of theory-based evaluations)*	Yes	2	2	2	4	3	4	5	5	3	5	5
1.4 Do you want to know "What other factors needed to be present alongside the intervention to produce outcomes observed? (Which factors were necessary and/ or sufficient for the intervention to work?)" *(Note this is a focus area of some evaluations where the intervention is not assumed to be the sole cause of change, but works in conjunction with other factors/ interventions)*	Yes	3	3	3	2	3	2	3	4	5	3	3
1.5 Do you want to know "Which outcomes of the intervention(s) being evaluated do different population groups consider to be the most important?" *(Note: this seeks to understand the relevance of the outcomes to different population groups or stakeholders)*	No	not applicable	not applicable	not applicable	not applicable	not applicable	not applicable	not applicable	not applicable	not applicable	not applicable	not applicable
STAGE 1 RESULT: Ability of a single evaluation method to answer all your key evaluation questions:		3.5	3.3	3.0	2.8	2.8	2.8	3.3	3.5	3.8	3.0	3.3

USERS: Look for a method that scores 5, or at least >4 for each individual evaluation question you want to answer (the coloured cells in rows 3-7). If you want to answer multiple questions, look at your overall "Stage 1 Result" (row 11). If there is a single method that scores 4 or more, that will be suitable; but - if as is likely if you want to answer multiple evaluation questions - there is no single method that scores 4 or more overall, consider combinations of methods that include at least one method that scores 4 or 5 for each individual question you want to answer. The appropriate combination will need to be discussed with evaluators.
Note that if you want to answer all 5 evaluation questions, it will require a combination of at least 3 of these methods to achieve a score of 4 or more against all individual questions.

STAGE 2: Features of Interest to Evaluation Commissioners or Managers that can Affect Choice of Methods

(note that some of these complement or elaborate upon key evaluation questions)

How desirable is it for you to be able to address each of the following areas of interest?	User Response	RCT (Randomised Contol Trial)	Difference-in-Difference	Statistical Matching	Outcome Mapping	Most Significant Change	Soft Systems Modelling	Causal Loop Diagram	Realist Evaluation	QCA (Qualitative Comparative Analysis)	Process Tracing/ Bayesian Updating	Contribution Analysis
2.1 I want to have groups of recipients and non-recipients who are perfectly similar except for their receipt of the intervention (Formally, eliminate selection bias between treatment and control groups complete and produce a perfectly unbiased estimation of the intervention effect)	Slightly desirable	33%	22%	22%	11%	11%	11%	11%	11%	11%	11%	11%
2.2 I want to allow all those meeting the criteria to participate in the intervention to actually receive it (i.e. prevent all eligible people from being excluded)	Desirable	22%	67%	67%	67%	45%	67%	67%	67%	67%	67%	67%
2.3 I want to be able to extrapolate or generalise the evaluation findings outside the cases or sample used for the analysis (i.e. external validity)	Desirable	45%	45%	45%	22%	22%	22%	45%	67%	67%	22%	45%
2.4 I want to allow the community (ies) in which the intervention was carried out to produce a collective evaluation of the most relevant changes at the community level	Not desired	not desired	not desired	not desired	not desired	not desired	not desired	not desired	not desired	not desired	not desired	not desired
2.5 I want to make a distinction between the achievement of minimum /expected goals and ideal / more ambitious programme goals	Slightly desirable	11%	11%	11%	33%	22%	11%	11%	11%	11%	22%	33%
2.6 I want to explore the higher order goals or values of the participants (like attitudes, norms, values and laws shaping their worldview, i.e. probing why certain results mattered to participants)	Not desired	not desired	not desired	not desired	not desired	not desired	not desired	not desired	not desired	not desired	not desired	not desired
2.7 I want the evaluation to capture a broad, systemic view of the situation (e.g. understanding how historical forces or path dependency or power relations or the economic system affect results, and seeing how those factors interact)	Slightly desirable	11%	11%	11%	11%	11%	33%	33%	22%	22%	22%	11%
2.8 I want the evaluation to make different perspectives about results or the causes of results explicit, particularly between programme participants representing different groups or households within the community, including the weakest	Slightly desirable	22%	22%	22%	22%	33%	33%	22%	22%	11%	11%	22%
2.9 I want the evaluation to identify and explain unintended changes and consequences, both positive and negative	Desirable	45%	45%	45%	22%	67%	67%	67%	67%	67%	67%	45%
2.10 I want to analyse complicated / complex mechanisms, including outcomes of non-linear relationships, vs. a mostly linear description of the programme Theory of Change	Desirable	22%	22%	22%	22%	22%	67%	67%	67%	45%	45%	22%
2.11 I want to obtain insights about the behaviour, attitudes and thinking of stakeholders	Slightly desirable	22%	22%	22%	22%	22%	33%	22%	33%	22%	33%	33%

2.12	I want to identify the different conditions that enable change in different contexts, as opposed to seeking a universal, population-wide or average explanation	Not desired	not desired	not desired	not desired	not desired	not desired	not desired	not desired	not desired	not desired	not desired	not desired
2.13	I want the evaluation to investigate what factors are necessary and / or sufficient for the intervention to produce results	Very Desirable	67%	67%	67%	33%	33%	67%	67%	67%	100%	100%	67%
2.14	I want the evaluation to measure confidence in one or more causal claims and for example determine whether the evaluation evidence is strong / conclusive for such claims or not	Very Desirable	100%	67%	67%	33%	33%	67%	67%	67%	67%	100%	67%
2.15	I want the evaluation to provide a detailed description of the process leading from programme activities to outputs, to intermediate outcomes and finally impacts	Very Desirable	33%	33%	33%	67%	33%	67%	67%	67%	33%	67%	100%
	Overall Score		59%	59%	59%	50%	48%	74%	74%	77%	71%	77%	71%
	# of Your "Very Desirable" Interests met in full	3	1	0	0	0	0	0	0	0	1	2	1

USERS: The method with the highest score is most able to address all your other interests.
However, consider also whether all of your "very desirable" interests are addressed by that method, by comparing the number in cell C23 with the numbers in cells Q23-AA23. If all your "very desirables" cannot be addressed by a single method, look for the combination of methods that could best address them, i.e. at least 1 dark green cell in those rows 4-18 which you have marked as "very desirable"

STAGE 3: How feasible is it to use each method to evaluate your intervention, given the requirements of those methods?

Note: this section asks about whether your intervention attributes 'fit' with the requirements for different evaluation methods to be used.
Qs3.1-3.10 ask about requirements for experimental and non-experimental methods; Qs3.11-3.19 ask about requirements for non-experimental or theory-based methods to work.
If your intervention cannot be conceived of in ways that fit with a particular method, that suggests the method will not be feasible - it is not a judgement about your intervention.

	(To what extent) Are you or the intervention to be evaluated able to meet the following conditions?	User Response	RCT (Randomised Contol Trial)	Difference-in-Difference	Statistical Matching	Outcome Mapping	Most Significant Change	Soft Systems Modelling	Causal Loop Diagram	Realist Evaluation	QCA (Qualitative Comparative Analysis)	Process Tracing/ Bayesian Updating	Contribution Analysis
3.1	To what extent was it (or will it be) possible for your intervention to control who participated in or received the intervention and who doesn't? *(Formally, what is your degree of control of treatment assignment?)* *(Note: if your intervention works with or targets an actor with no comparable non-recipient - e.g. advocacy towards the sole actor working on an issue - please choose 'not at all')*	Not at all	0%	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method
3.2	How many individuals or households / villages / geographic regions etc. are participating in or receiving the intervention? *(Formally, how large is the treatment group?)* *[Note: if your intervention is focused on advocacy/ policy influencing, consider each advocacy target]*	<30	0%	0%	0%	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method
3.3	How many individuals or households / villages / geographic regions etc. are in a similar situation to the participants but do not participate in the intervention? *(Formally, how large is the control group?)* *[Note if your intervention does not and could not have a control group, please answer '<30']*	<30	0%	0%	0%	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method
3.4	To what extent are participants in the intervention isolated from non-participants? *(Fully = isolated; not at all = they communicate and interact to such an extent that the intervention is highly likely to affect outcomes observed in non-recipients)* *[Note if your intervention does not and could not have 'non-participants' or a control group, please answer 'not at all']*	Not at all	0%	0%	0%	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method
3.5	To what extent are intervention participants and non-participants (or treatment and control groups) subject to the same external influences for the entire duration of the intervention? *[Note if your intervention does not and could not have 'non-participants' or a control group, please answer 'not at all']*	Not at all	0%	0%	0%	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method
3.6	Are some potential participants who meet eligibility criteria to receive or participate in the intervention excluded from it? (e.g. when the intervention is rationed) *[Note: if this is not applicable to your intervention, please answer 'not at all']*	No	0%	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method

(Left vertical label: or quasi-experimental evaluation methods to evaluate your intervention?)

#	Question	Answer	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
3.7	Is information on the background characteristics of those participating and not participating in the intervention gathered using identical or highly comparable methods? *[Note: if this is not applicable to your intervention, please answer 'not at all']*	No	0%	0%	0%	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method
3.8	Can at least 30 pairs or couples of participants and non-participants be identified and "matched" on the basis of the above-mentioned high quality data or comparable surveys? *[Note: if this is not applicable to your intervention, please answer 'no']*	No	Does not affect ability to use this method	Does not affect ability to use this method	0%	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method
3.9	To what extent is high-quality baseline data available for both those participating and not participating in the intervention (treatment and control groups)? *(Note baseline data refers to indicators on intended outcomes and outputs, measured prior to the start of the intervention. It is distinct from "background characteristics" above. If this is not applicable to your intervention, please answer 'not at all')*	Not at all	0%	0%	0%	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method
3.10	To what extent has previous (time series) data confirmed that the trend in your outcome indicators of interest before the intervention started was the same for both those receiving and not receiving the intervention (treatment and control group)? *(Formally: is the common trend assumption verified?)* *[Note: if this is not applicable to your intervention, please answer 'not at all']*	To some extent	Does not affect ability to use this method	67%	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method
3.11	To what extent is information on (at least a small number of) factors which are assumed to affect the outcome consistently available across at least 5 or 10 cases? *(Note: each 'case' can refer to an application of the intervention in different locations/ contexts, or among different individuals, institutions or groups. If the concept of cases is not applicable to your intervention, please answer 'not at all')*	To some extent	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	67%	Does not affect ability to use this method	Does not affect ability to use this method
3.12	To what extent are excellent facilitation skills available in the evaluation team? *(Note: if your evaluation team has not yet been appointed, please answer based on skills you will be seeking and your expected ability to secure those. This question highlights those methods for which such skills are most important)*	Don't know	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	not known	not known	not known	not known	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method
3.13	To what extent does the success of the intervention depend on the behaviour of several stakeholders which may be relatively unpredictable? *(Note: this question highlights methods particularly suited to working within complex systems, e.g. where you are trying to affect social dynamics or market dynamics)*	Poorly	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	33%	Does not affect ability to use this method	17%	33%	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method
3.14	To what extent are both those likely to conduct the evaluation and those who will be consulted for the evaluation likely to be open to airing different perspectives on the intervention, its outcomes and how change happened? *(e.g. the power dynamics are such that multiple worldviews could be expressed, rather than only one dominant worldview being voiced)*	Don't know	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	not known	not known	not known	not known	not known	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method

Is it feasible to use experimental

...n methods to evaluate your intervention?

Is it feasible to use non-experimental or theory-based evaluation

#	Question	Extent	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
3.15	When examining a number of different cases, to what extent do you expect the evaluation team to be able consistently to get an understanding of the contextual factors that affected the outcomes of your intervention? *(E.g. if your intervention deals with different locations, population groups or institutions that can affect how the mechanisms between your intervention and outcomes work)*	Fully	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	100%	100%	100%	50%	50%	50%
3.16	To what extent is the evaluation team be able to formulate, test and refine theoretical assumptions about the behaviour, attitudes and thinking of stakeholders? *(i.e. "identifying the mechanisms generating the outcomes"; this might require good insights in political science, psychology, social sciences, or specific domains; hence might be easier in multi-disciplinary teams)*	Fully	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	100%	50%	50%	50%
3.17	To what extent is the evaluation team be able to map or understand complicated / complex mechanisms, including outcomes of non-linear relationships? *(Note: this is as opposed to taking a mostly linear description of the programme Theory of Change or intevention logic)*	Fully	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	100%	100%	50%	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method
3.18	To what extent are evaluators able to access to a broad range of detailed and high quality data necessary to answer your evaluation questions, including hard to find data? *(Note: this could include - for example - data on sensitive issues (e.g. on protection/ safeguarding issues); or data from conflict-affected or hard-to-reach areas/ populations; or minutes of private meetings, personal emails,*	To some extent	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	34%	34%	Does not affect ability to use this method	Does not affect ability to use this method	67%	Does not affect ability to use this method
3.19	To what extent can you be confident that your chosen evaluator is able to set up a Theory of Change with a causal chain, and risks and assumptions for each step that would help shape complementary or alternative explanations for observed changes, either from scratch or making use of an existing theory of change for the intervention?	Fully	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	Does not affect ability to use this method	50%	50%	50%	100%
Number of Essential Requirements to Use this Method that your intervention cannot meet			6	6	6	0	0	0	0	0	0	0	0
Number of Essential Requirements to Use this Method which you "don't know" if you can meet			0	0	0	0	1	2	1	0	0	0	0
Number of Desirable Requirements to Use this Method that your intervention cannot meet			2	0	1	0	0	0	0	0	0	0	0
Score as % of Maximum for that Method			0%	10%	0%	22%	0%	50%	53%	75%	87%	87%	100%

USERS: Look for the evaluation method(s) with either zero or the fewest "essential requirements" unmet: this should be the most feasible method to use given your intervention attributes. If there are any essential requirements to use a particular method that your intervention cannot meet, you will have to either address that constraint (if possible), or choose another method that may answer your key evaluation questions.

If you don't know whether an essential requirement for an evaluation method can be met - and if that method otherwise seems appropriate - try to get an answer to any "don't knows" before making a final choice. "Desirable" requirements are less binding, but worth considering.

Combining Stage 1 and Stage 3: Are the evaluation methods that are most feasible to use given your programme attributes (Stage 3) the same as those that are most suited to answering the evaluation questions you are interested in (Stage 1)?

If not, consider whether any changes can be made to your intervention to address any 'essential requirements not met', or - if that cannot be changed - try different evaluation questions to see which can best be answered with the methods that are feasible to use given your intervention attributes.

SUMMARY RESULTS - ALL STAGES	RCT (Randomised Contol Trial)	Difference-in-Difference	Statistical Matching	Outcome Mapping	Most Significant Change	Soft Systems Modelling	Causal Loop Diagram	Realist Evaluation	QCA (Qualitative Comparative Analysis)	Process Tracing/ Bayesian Updating	Contribution Analysis
Stage 1: Which Method is Best Suited to Answering My Key Evaluation Question(s)?	3.5	3.3	3.0	2.8	2.8	2.8	3.3	3.5	3.8	3.0	3.3
Stage 2: Which method is most able to address my other interests?	59%	59%	59%	50%	48%	74%	74%	77%	71%	77%	71%
Stage 3: Which Method has the fewest essential methodological requirements that cannot be met by my intervention? (Which method is most feasible to use?)	6	6	6	0	0	0	0	0	0	0	0

USERS: consider the compatibility of the results across the three stages.
For example, is there a single tool that is well suited to answering your evaluations questions (scores 4 or more in Row 2), can address your other interests (Row 3) and is also feasible given your intervention attributes (Row 4)?
There rarely will be a single appropriate method. Users will typically need to either look for a combination of methods that provides a "best fit", or - if there is flexibility - users may want to revisit the evaluation questions, interests and intervention attributes to see if any of these parameters can be changed to make some methods more appropriate.
HEALTH WARNING! This tool should only be used as an aid to discussion between evaluators and commissioners, especially when combinations of methods are required; it is not intended as a basis for decision-making on its own.

If there are significant differences between your results for the most appropriate methods at each stage, options for finding a solution include:
- Review results for each stage in detail to see if a combination of methods could be both appropriate and feasible
- Engage in dialogue with intended users or other stakeholders to reconcile the intended scope of the evaluation with the intervention attributes. Try narrowing the focus of the evaluation.
- If your intervention/evaluation attributes are not yet fixed, consider altering the intervention/evaluation to make it feasible to answer all your key evaluation questions while also meeting all essential requirements of one or more evaluation methods