



Article

Rethinking rigour to embrace complexity in peacebuilding evaluation

Evaluation

1–26

© The Author(s) 2024



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: [10.1177/13563890241232405](https://doi.org/10.1177/13563890241232405)

journals.sagepub.com/home/evi



Marina Apgar 

Institute of Development Studies, UK

Helene Bradburn

Independent, France

Livia Rohrbach

Search for Common Ground, Tanzania

Leslie Wingender

Humanity United, USA

Edwin Cubillos Rodriguez

Everyday Peace Indicators, Colombia

Angela Baez-Silva Arias

Adapt Peacebuilding, Colombia

Alamoussa Dioma

Malian Institute of Action Research for Peace, Mali

Tiffany Fairey

King's College London, UK

Stephen Gray

Adapt Peacebuilding, Spain

Corresponding author:

Marina Apgar, Institute of Development Studies, Library Road, Falmer, East Sussex BN1 9RE, UK.

Email: m.apgar@ids.ac.uk

Ayak Chol Deng Alak

Inclusive Peace, South Sudan

Steff Deprez

Voices That Count, Belgium

Abstract

The field of peacebuilding evaluation has evolved over time in response to the complex nature of peace efforts. However, it still predominantly relies on evaluation models that aim to measure discrete peace outcomes adhering to rigid notions of rigour. The inclusive rigour framework presented in this article responds to this challenge, adding to complexity-aware and epistemologically plural approaches to build credible causal explanations in conditions of uncertainty. It identifies three interconnected domains of evaluation design and practice: effective methodological bricolage, meaningful participation and inclusion, utilisation and impact. Rigour here is not defined by methodological choice alone, but rather, relies on an active view of evolving methodological choices throughout an iterative process as maximum use value and meaningful participation are sought. Using three cases, we highlight the critical role of partnership arrangements and associated evaluation cultures and mindsets underpinned by power dynamics that enable or hinder the practice of inclusive rigour.

Keywords

bricolage, complexity, participation, peacebuilding, rigour

Introduction

Navigating the multifaceted challenges of evaluation in complex settings, particularly in the realm of peacebuilding, requires approaches that go beyond traditional, standardised evaluation models. The field of peacebuilding evaluation has evolved over time in response to the complex nature of peace efforts. However, it still predominantly relies on evaluation models that aim to measure discrete peace outcomes and adhere to rigid notions of rigour. This presents a fundamental challenge as there is no consensus on the definition of ‘peace’ nor how to effectively measure it. The absence of violence, often associated with negative peace, does not fully encompass the concept of positive peace, which includes the broader set of structures, institutions, behaviours and attitudes that sustain safe, healthy, just and peaceful societies. The ongoing evolution of definitions further complicates the matter with more agreement on conflict definitions than on those of peace itself (Firchow, 2018; Gleditsch et al., 2014).

Peacebuilding interventions exist within intricate, dynamic and potentially volatile contexts, characterised by systemic relationships involving multiple actors. These contexts do not follow stable trajectories and frequently undergo unpredictable changes in their causal landscapes. Moreover, peacebuilding is inherently political, influenced by factors of both local and international origin. This complexity makes it challenging to discern, predict or measure the impact of peacebuilding efforts. The unpredictable and highly relational nature of these causal pathways renders most attempts to measure predefined outcomes ineffective, highlighting the need for a more nuanced and learning-oriented approach, accompanied by new conceptions of rigour.

Calls to reframe rigour to address complexity and foster inclusion are not novel. Inclusive rigour (Chambers, 2015) and adaptive rigour (Preskill and Lynn, 2016) have been discussed, with further refinements by Aston et al. (2022) and Aston and Apgar (2022) in the context of complexity-aware evaluation. In this article, we present a framework for inclusive rigour that builds on these advances and places participation at its core, considers appropriate methodological choices and responds to the complexity inherent in peacebuilding interventions and their evaluation. Rigour here is not defined by methodological choice alone, but rather relies on an active view of evolving methodological choices that unfold throughout an iterative process as maximum use value and meaningful participation are sought. The framework has been collaboratively developed over several years through reflection and learning within a community of practice comprising development and peacebuilding evaluators, researchers and implementers.

Evaluation in complex settings

Peacebuilding interventions, set within complex and dynamic contexts, present unique challenges for evaluation. These environments, characterised by unpredictable and often volatile change and multiple political influences, with marked power imbalances, often part of colonial legacies and embedded social norms, complicate the discernment, prediction and measurement of outcomes. The journey of peacebuilding evaluation has witnessed a transformation over time, grappling with the complex nature of peace efforts (Delgado et al., 2022). Oakley (2022) discusses the challenges posed by complexity in the evaluation of democracy, human rights and governance programmes more broadly, underscoring the need for evaluations to be adaptable and sensitive to the multifaceted nature of political and social landscapes. Despite this evolution, the field continues to cling to hierarchical models and rigid notions of rigour (Fairley and Kerr, 2020; Pearson d'Estree, 2020), often overlooking the dynamic nature of peacebuilding (Scharbatke-Church, 2011). Standardisation of evaluation models continues to be the norm, evidenced by Bornstein's (2010) Peace and Conflict Impact Assessment (PCIA) and as shown by Paffenholz (2015) and Andersen et al. (2013). This phenomenon creates a 'rigour trap', where established methods overshadow the adaptive nature of peacebuilding evaluation and its interconnected utility posing a critical challenge.

It is now widely recognised that incorporating local, lived experiences of conflict-affected communities is essential for a comprehensive assessment of peacebuilding interventions. Paffenholz (2015) and Firchow (2018) advocate for a peace that originates 'from below', firmly grounded in the community's perceptions of peace. The Grounded Accountability Model (GAM), as recently discussed by Urwin et al. (2023), exemplifies such an approach, illustrating how accountability can be co-constructed between external actors and the communities in conflict-affected settings, thereby enhancing localisation and programme effectiveness. Participation, therefore, emerges as a pivotal theme within the evolution of peacebuilding evaluation. Notably, there is progress in participatory methodologies, with an emphasis on meaningful participation highlighted by the work of Firchow and Selim (2022), who prioritise meaningful participation over mere quantitative involvement. A significant avenue that has garnered attention is the participatory formulation of indicators in collaboration with local stakeholders (Bornstein, 2010; Firchow, 2018; Urwin et al., 2023). This practice revolves around involving those directly impacted by peacebuilding efforts in the evaluation process, enhancing the contextual relevance and sensitivity of assessments. However, a critical question remains within these advancements: do current participatory

approaches effectively encompass both complexity awareness and the capacity for meaningful causal inference (Firchow and Selim, 2022)?

To address the challenges outlined, some have proposed broadening evaluation approaches to embrace complexity-aware frameworks, akin to developments in the broader field of evaluation of international development programming (Befani et al., 2015). The adoption of these approaches, together with the attention to participation, could provide a means to overcoming the ‘rigour trap’. In the following section, we review the evolution of the concept of ‘rigour’ in evaluation theory and practice, identifying relevant trends related to evaluation in complex settings. Subsequently, we introduce the framework, demonstrating alignment with some of the emerging trends. We then share our learning from applying the framework across three distinct experiences of peacebuilding evaluation.

The contested terrain of ‘rigour’ in evaluation

Our starting point for exploring the evolution of the concept of ‘rigour’ in the broader field of evaluation is an appreciation of evaluation as applied social research that at its core applies evaluative reasoning (Davidson, 2013) to build understanding of the operation and outcomes of interventions. It is about ‘valuing’ what is being achieved and how programmes work in order to inform future programming and funding decisions (Schwandt and Gates, 2021). Focusing on the ‘use value’ of the results of evaluation places attention on the quality of evaluative judgements, or what is known as its ‘probative value’ (Ribeiro, 2019). This remit of evaluation has led to a long and vibrant discussion around appropriate frameworks for assessing ‘quality’ in evaluation (Downes and Gullickson, 2022).

Criteria used to assess quality in evaluation have largely been borrowed from social sciences, which, given their academic origins are less concerned with use and more concerned with disciplinary oriented framings of quality. As White (2019) categorises in his ‘four waves of evidence’ the advance of the ‘what works’ agenda has dominated frameworks based on knowledge and evidence hierarchies stemming from positivist epistemologies and research designs from the medical sciences. They typically have experimental designs (with randomised control trials as the gold standard) at the top. Even if we know that experimental designs are not always appropriate, relevant, feasible, or ethical (National Institute for Health and Care Excellence (NICE), 2012; Stern et al., 2012) particularly in the context of evaluating changes in complex social phenomena, the primacy of knowledge hierarchies still influences mainstream framings of quality, validity and rigour in evaluation.

As Downes and Gullickson (2022) show, conceptualisations of ‘validity’ are contested, with over 40 different ways of interpreting quality in evaluation. When the term ‘validity’ was initially used in evaluation, it was focused primarily on the method employed. The Campbellian validity framework (see Campbell and Stanley, 1963) was based on validity criteria appropriate for quantitative methods, and remains dominant today. Evaluators often use its four forms of validity (internal, external, construct and statistical inference) as a guiding framework (see Jiménez-Buedo and Russo, 2021). Even when mixing of quantitative and qualitative methods is proposed, these four forms of validity still frame conceptualisations of rigour (Maxwell, 2004; Ton, 2012).

This dominant approach has been critiqued, leading to an evolution of frameworks within the social sciences and by extension evaluation theory. Building on constructivist criteria (Guba and Lincoln, 1989), which are framed around the concept of ‘trustworthiness’ criteria such as credibility, transferability, dependability and confirmability today are seen to enhance the Campbellian view of validity (Bamberger et al., 2010).

Going a step further, some argue that reasoning and judgement in evaluation should be the main driver of validity, rather than a focus on measurement and methods (Hurteau and Williams, 2014; Scriven, 1995). As House (2014) puts it, there are ‘many circumstances in which the arguments for validity via technical adequacy fall short’ (p. 13). He proposes more useful standards to assess how reasoned, fair and convincing evaluative arguments are. This view of validity moves from a methods-centric focus to embrace the relationship between the ‘probative’ and ‘use’ value of evaluation.

In the context of adaptive programmes that intentionally respond to complexity and acknowledge political context as a key factor in evaluation, the ‘use value’ ascribed to evaluation creates demand for ‘actionable evidence’ (Pasanen and Barnett, 2019). Notions of ‘adaptive rigour’ (Preskill and Lynn, 2016) have recently been evolved in response to complexity-aware evaluation, which aims to harness such actionable learning (Aston et al., 2022). They offer the following integrated set of criteria: reasoning, credibility, responsiveness, utilisation and transferability. Together, they provide a new landscape for evaluators and programmers working in conditions of complexity and seeking to build credible evidence in ways that respond to demands of multiple stakeholders, including those that tend to be excluded or marginalised. The responsiveness criterion in particular opens the door for moving beyond evaluation as simply a technical endeavour to embrace it as a political process involving diverse stakeholders who may not all see eye to eye (Apgar and Allen, 2021; Roche and Kelly, 2012).

The concept and practice of rigour is likely to remain contested within the evaluation field, reflective of the plurality of evaluation theory and practice. In the context of peacebuilding evaluation, there is still a need to broaden beyond simple knowledge hierarchies that continue to inform rigid views of rigour. The evolution we have outlined above related to evaluation that can work with rather than against complexity is a useful starting point, creating the opportunity the field needs to build greater congruence between the realities of peacebuilding interventions and their evaluation. In this article, we progress this evolution by developing an integrated framework of ‘inclusive rigour’ as one such alternative.

Co-developing an integrated framework for inclusive rigour

As peacebuilding monitoring, evaluation and learning (MEL) practitioners and researchers, we have found ourselves needing to push beyond existing frameworks and methodological approaches commonly applied. We came together as a community of practice in 2020 to learn as we put complexity-aware and participatory evaluation into practice, holding the question of rigour as a central concern across evaluation design and the conditions that enable or hinder it. The framework we present here is the result of an intentional facilitated learning process across the co-authors and members of the community of practice, using an ‘action science’ orientation (Friedman, 2008).

Friedman describes the process of ‘action science’ as ‘creating a community of inquiry within a community of practice and building theory through combining testing and practice with rigorous interpretation’ (Friedman, 2008: 11). Our process has included internal moments of reflection and learning, periods of application and testing through our empirical work in the field, which in turn supported external moments of sharing, reflection and learning with our broader communities of practice through engagement in three conferences (EES 2021 online, PeaceCon@10 online 2022, EES 2022 in person).

We first grounded ourselves in sharing our different experiences of grappling with evaluation designs to enable participation and rigour, which we discussed with the broader evaluation community at the European Evaluation Society (EES) conference online in

Table 1. Literature on rigour in evaluation that guided learning from practice.

Elements of rigour	Learning questions for exploration of rigour through our cases
From Chambers, 2015 – Canons of Inclusive Rigour for Complexity	
Canon 1. Eclectic Methodological Pluralism	How do we combine and recombine methods for ‘good fit’, to ensure to optimise learning and adaptation?
Canon 4. Triangulation	
Canon 3. Adaptive iteration	What mechanisms/processes enable ongoing reflection and iteration of methods to enable ongoing ‘good fit’?
Canon 2. Improvisation and Innovation	What conditions enable improvisation (bricolage) and innovation?
Canon 5. Plural perspectives	How do we optimise participation and inclusion through our methods?
Canon 6. Optimal ignorance and appropriate imprecision	What ethical standards do we use to ensure we do not overburden communities?
Canon 7. Being open alert and inquisitive	What personal and team competencies support inclusive rigour?
Transparent reflexivity	What institutional arrangements enable reflexive teams?
From Preskill and Lynn (2016) and Aston et al. (2022)	
Reasoning	How does our combination of methods and processes support critical reflection on how change happens?
Credibility	Whose confidence in findings matter? How do our methods ensure internal validity leading to high degree of confidence?
Responsiveness	How do our methods and processes ensure plural perspectives and respond to local values and needs?
Utilisation	How do our methods and processes ensure learning responds to multiple stakeholder demands?
Transferability	How do we include context within methods and processes to enable learning that is useful beyond the specific project?

Source. authors own.

2020. Having identified rigour as our central focus of co-inquiry, we then conducted a full literature review of existing frameworks that could serve to further structure our learning from practice (see Table 1). We focused first on the canons of inclusive rigour from Chambers (2015) and then explored the approaches put forward by Preskill and Lynn (2016) and Aston et al. (2022) with their comprehensive criteria of adaptive rigour. The similarities across the three frameworks were useful starting points for our inquiry and the focus on methodological mixing, attention to plural perspectives and a pragmatic focus on utilisation led us to define more specific learning questions to explore rigour through our experiences (shown in Table 1).

We created an initial framework of ‘inclusive rigour’ that informed our workshop at PeaceCon 10, which looked across our experiences to explore if an integrated framework was useful. We then used our critical reflections from this event to further explore the framework as a means to intentionally learn across our experiences, approaching them as experiments in the practice of inclusive rigour. We further evolved the framework to better specify the conditions that enable or hinder inclusive rigour to take shape in practice. At the 2022 EES Conference, we held a session that elaborated on the methodological bricolage aspects of our evaluation designs, then returned to our practice and deepened our learning across the case studies. In the following section, we present the resulting, evolved framework.

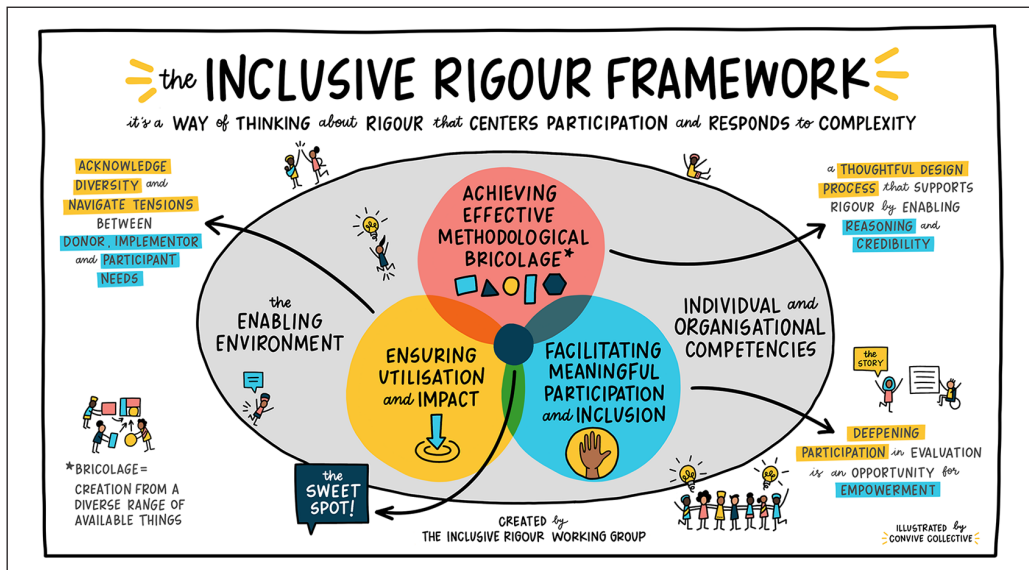


Figure 1. The inclusive rigour framework.

Introducing the inclusive rigour framework

The framework is shown in a visual form in Figure 1, and suggests that inclusive rigour in peacebuilding evaluation becomes operational through three interconnected domains of design and practice. Each of the domains has been described in evaluation literature already and will not be unfamiliar to evaluation practitioners. In our description of these domains as core elements of designing for and practising inclusive rigour, we illustrate the theory and practice we are building on and highlight our specific interpretation.

Achieving effective methodological bricolage

The domain of methodological bricolage includes negotiating and making decisions about appropriate design, methodological mixing and choices of specific methods and tools we bring together for the purpose of understanding and evaluating causal pathways. The term methodological bricolage has a long history in anthropology, first coined by Lévi-Strauss (1966) and is generally understood as the practice of using a heterogeneous repertoire of available tools to solve new problems. This resonates with the pragmatic design choices and creativity often required by evaluators as they build fit for purpose designs to respond to multiple user needs in ever evolving contexts. Yet it is only recently gaining ground as a concept in evaluation and specifically within complexity-aware and systemic approaches, which emphasise learning, use and multiple forms of knowledge (Hargreaves, 2021; Patton, 2019).

Aston and Apgar (2022) describe it as

Evaluators often only adopt certain parts of methods, and skip or substitute recommended steps to suit their purposes. The evaluator may repurpose existing tools with those of methods and tools with which they are more familiar; or they may even combine a patchwork of relevant tools for different parts of an evaluation or throughout the cycle of designing, planning, monitoring, and evaluating a project (p. 2).

Chambers (2015) referred to this as ‘eclectic methodological pluralism’ agreeing that methods chosen in evaluation must be a ‘good fit’ for the context and the users. In his view, rigour stems from ‘scanning the range of possibilities and adapting and combining these for the special conditions of each inquiry’ (Chambers, 2015: p. 328). Van Hemelrijck and Guijt (2016) then added the need to balance inclusiveness, rigour and feasibility when mixing methods. Similarly, Aston and Apgar (2022) suggest that quality in bricolage rests on careful consideration of what function a particular method or part of a method serves in the evaluation process, and how this supports rigour by enabling reasoning, credibility and transferability.

Bricolage as an intentional approach to evaluation design extends long-standing discussions about methodological choice that have led to a proliferation of support tools, such as the design (HM Treasury, 2020; Stern et al., 2012) and choice (Befani, 2020) triangles, which suggest choice should be guided by the evaluation questions, the programme attributes and the intended use. Central to this view of methodological selection is consideration for the underlying, and often hidden, frameworks of causal inference and how these relate to specific evaluation questions (see Gates and Dyson, 2017; Jenal and Liesner, 2017; Lynn and Apgar, 2024). Asking ‘how’, ‘why’ and for ‘whom’ questions, which are increasingly common in evaluation of international development and peacebuilding interventions, require methods that use ‘generative’ and ‘configurational’ approaches to causality, in other words, that acknowledge multiple interacting factors understood to work in specific ways in any particular context. It is, therefore, within the methodological bricolage domain of the framework, where evaluation questions are considered, that we recognise the need for design choices that will enable appropriate causal analysis through sound reasoning (including looking for alternative explanations) leading to credible evidence (with high confidence in findings and causal claims) and greater potential for transferability (exploration of the role of context within the causal claims).

Practising methodological bricolage in peacebuilding evaluation can contribute to building rigour not through choosing the ‘right’ most ‘rigorous’ methodological design based on a hierarchy of knowledge, but rather through combining different methods or parts of methods as they enable credible evidence (containing strong and justified causal claims) to be generated along unpredictable causal pathways in response to context and stakeholder needs.

Facilitating meaningful participation and inclusion

The domain of meaningful participation and inclusivity is how we pay attention to the ways in which our processes open up or close down space for different forms of knowledge, particularly of the most marginalised, to be included meaningfully. A number of frameworks examine the purpose and forms of participation of different stakeholders in evaluation. Some place participation along a spectrum ranging from ‘instrumental’ to ‘emancipatory’ (Cousins and Whitmore, 1998). Others take dichotomous views of participation as ‘instrumental’ versus ‘normative’ (Baker and Chapin, 2018) or ‘technocratic’ versus ‘participatory’ (Chouinard, 2013). Instrumental/technocratic reasons often relate to utilisation, with participation of a range of stakeholders seen as enhancing the possibility of uptake and use. Instrumental reasons can also rest on acknowledging that causal explanations of how change happens are not value free, and triangulation across diverse experiences can support more credible causal claims, particularly when dealing with change in complex systems.

Normative or emancipatory approaches see evaluation as a process aiming to support transformative change. This includes empowerment (Fetterman, 1994) and feminist (Brisolara et al., 2014) evaluation, which explicitly support social justice agendas and are concerned with questions of power. While these heuristic tools to define forms of participation are overly simplistic, they usefully highlight a central concern around what constitutes meaningful participation within evaluation: the need to pay attention to both ‘who’ is participating, as well as ‘how’, in order to establish at what ‘depth’ participation is desired and appropriate, to design the participatory process accordingly.

What we are concerned with in this domain of the inclusive rigour framework, is the desire to move practice towards ‘deeper’ forms of participation paying particular attention to processes that enable excluded populations to engage meaningfully. In our context of peacebuilding programming, there are considered attempts to practise more locally led and decolonised evaluation (Chilisa and Mertens, 2021; Forsyth et al., 2021; Kelly and Htwe, 2023). While we agree with critics that ‘decolonising’ is at risk of becoming an empty buzzword and should not be used metaphorically (Tuck and Yang, 2012), we see this opening towards more radical forms of participation within evaluation practice as an opportunity for movement towards inclusive rigour, away from a simple hierarchy of evidence, and centering the local voices concerned with the evaluand. For example, calls for a fifth branch to the tree of evaluation approaches that highlight ‘context and needs’ (Chilisa, 2019) brings the ontological, epistemological and axiological foundations of indigenous knowledge systems, which have historically been excluded, into the evaluation picture as a starting proposition and not solely as a methodological conundrum. This allows us to tackle both the responsiveness and transferability criteria head on.

This evolution of evaluation that is driven in and by context and by those closest to the experiences of change, brings necessary precision to what has been a central theme in participatory evaluation: being aware of and engaging with power dynamics makes the difference between instrumental and transformative forms of practice. Hanberger (2022) offers a useful framework through which to explore power ‘in’ and ‘of’ evaluation, illustrating that it operates both at the methodological and practice levels, in the doing of the evaluation, as well as at the governance level where the power of evaluation lies in how it is used by stakeholders. We reflect further on the power ‘of’ evaluation as we describe the ‘enabling environment’ for meaningful participation ‘in’ the evaluation process itself.

Ensuring utilisation and impact

The domain of utilisation is where we strive to respond to different stakeholder needs for evidence and learning to inform decision-making and achieve the ultimate goal of increasing the impact of peacebuilding programmes on the ground. The evaluation field has been discussing ways to support and overcome barriers to greater ‘use’ of evaluation for decades, with recent reviews (Pattyn and Bouterse, 2020) illustrating a multitude of factors at play across individual, organisational and system dynamics. Two areas are increasingly seen as important and relate to our framing of this domain – stakeholder involvement and evaluator competence (Johnson et al., 2009). We will explore the latter in the following section on the enabling environment.

There are a range of potential stakeholders who can support uptake and use. The list includes: (1) funders or commissioners who use findings to inform strategy; (2) programme

staff, who, especially with learning-oriented evaluations, become the main users of the findings, and whose proximity to the theories of change and action suggest they can support appropriate evaluation design; (3) intermediaries and partners (such as non-governmental organisations (NGOs) or service providers) who the programme engages with as a way of reaching primary beneficiaries, and who are key players in the processes of change being examined; and (4) the people whom the programme being evaluated is aiming to serve who most directly experience the impact. These roles are often not static and can change throughout an evaluation process and will depend on the specific evaluation parameters. And as meaningful participation is deepening to support local demands, we can expect the overlapping roles to evolve further.

Ensuring utilisation requires not just acknowledging the different positions, but crucially navigating across, between and through them. In this vein, some argue that a focus on the political challenge is central to impactful evaluation practice (Aston et al., 2022; Roche and Kelly, 2012). Similar to scholarship on the politics of evidence use (Parkhurst, 2017), it is naive to assume that decision-making around use of the evidence produced through evaluation is merely a technical or scientific endeavour. Utilisation focused and responsive evaluation approaches focus precisely on this aim of balancing power asymmetries that might arise within the evaluation process and that directly influence use (Baur et al., 2010).

Stakeholders bring their different forms of power to influence the methodological choices made (domain of methodological bricolage) as well as the extent to which marginalised communities are included meaningfully (domain of inclusivity). All stakeholders can potentially influence the extent to which an evaluation design and its implementation is credible as well as whether the design allows for context to be considered for transferability of the results. Local stakeholders, for example, may not prioritise transferability over their immediate use in context, while a commissioner may wish to learn how to apply similar strategies in other contexts, placing transferability at the top of the agenda. Application of certain preferred methods may be necessary to have convincing levels of credibility in the eyes of some stakeholders while responding to local or partners' needs may call for greater methodological innovation and flexibility.

What becomes important in this domain, therefore, is to pay particular attention to the quality of the governance processes that can enable a deliberative space for learning across different stakeholders. In the context of peacebuilding evaluation, in conditions of complexity, ensuring the findings are used to adapt programme implementation on the go, places particular emphasis on the needs of programme implementers perhaps beyond the needs of external actors. And negotiation does not always lead to consensus, suggesting that thinking about 'use' in the context of inclusive rigour may require hard choices to be made to ensure maximum use where it can bring maximum impact.

Enabling environment and organisational and individual competencies

We have described the three interconnected domains of practice to operationalise inclusive rigour. And as described across all three domains, individual peacebuilding evaluators and evaluation teams are not working in isolation. They are part of structures, institutional arrangements, broader systems of aid, evaluation and evidence use and interact within both formal and informal spaces that together enable or hinder inclusive rigour. We highlight two salient aspects of this broader enabling environment and expand on each in turn.

Institutional dimensions

The power 'of' evaluation to support all domains of practice and inclusive rigour is conceptualised through the structures and policies that govern evaluation (Hanberger, 2022). Much of the discussion about the politics of evidence in international development (see Eyben et al., 2015) has centred on the hierarchies of aid and the flow of resources from funders down to implementers, which drives the need for upwards accountability and the most difficult power imbalance to shift. The role of funders remains critical to supporting methodological bricolage, participation and opening up to plural views of use as much peacebuilding is part of formal funded development interventions. We do see in certain funding circles a much greater appreciation for funding in ways that navigate power asymmetries, shift power and emphasise new ways to partner (Gibson, 2018; Trust Based Philanthropy Project, 2023). Frameworks that explore equitable partnerships in the context of international development argue that appreciating historical context, which is colonial and driven from the Global-North and comes with associated power asymmetries, should be the starting point (Fransman et al., 2021; Snijder et al., 2023). Equitability is characterised by joint ownership, mutual responsibility, transparency and benefit sharing for all partners (Price et al., 2021). Where these dimensions are made explicit and inform the governance arrangements of evaluation, the conditions will be more favourable for inclusive rigour.

But even when donors are shifting their models of partnership, the historical legacy of top-down aid shaping most monitoring and evaluation as an instrument of control around predefined results, manifests today in the cultures and mindsets that drive institutional dynamics of partners. Throughout the development and peacebuilding systems across scales and institutions at both the local and national levels, the command-and-control management practices and associated cultures they are part of continue to be perpetuated daily. Shifting these cultures towards learning is at the heart of enabling effective methodological bricolage. This includes managing uncertainty and moving away from a control-orientation to create space, time and budget for flexibility and iterative co-design. Those supporting adaptive management make the case that the enabling environment requires shifts across not just evaluation but other institutional domains such as contracting and compliance that can often become barriers to broadening the risk landscape and embrace flexibility (Prieto-Martín et al., 2017).

Personal and team competencies

Alongside the structures and institutional cultures lie the competencies required to practise in all three domains. A set of relational and political competencies support evaluators to act as facilitators of learning. As Eager and Barnett (2021) show, ethics and roles of evaluators when working in conditions of complexity shift away from assessing impact as independent agents to becoming an integral part of achieving impact as embedded evaluators.

The competencies required to make this shift have been described within the context of participatory evaluation and include: sound facilitation skills and reflexivity; humility and honesty; balancing principles with pragmatism and understanding the political landscape (Apgar and Allen, 2021; Podems, 2010). Understanding the political landscape enables navigating multiple competing stakeholder demands (utilisation and impact domain) and can help evaluators to decide when to push for particular methodological combinations (methodological bricolage domain), and when to deepen participation (domain of meaningful participation). Perhaps

the core competency that underpins quality in bricolage is being able to balance principles with pragmatism – evaluation is never an exact science. Chambers' (2015) canon of 'being open, alert and inquisitive' and 'employing transparent reflexivity' links to a call for greater humility and honesty in embedded and facilitative evaluations. Feminist evaluators would take this even further, to argue that evaluators who are part of the process must recognise what they are bringing into that process (Patton, 2002; Podems, 2010), seeing themselves as advocates and facilitators of processes aimed at empowerment (Miller and Haylock, 2014).

These individual competencies are enabled and supported through team and institutional competencies. In the adaptive management field, it is acknowledged that 'a culture and mindset that encourages and rewards open, alert, inquisitive, anticipatory, responsive and honest approaches' (Ramalingam et al., 2017, 2019) builds a conducive enabling environment. These competencies can be built intentionally, but often are not all available at the outset. This raises important questions about how to balance the need for independence and expertise that remain core attributes of evaluators, while also enabling learning and navigating different stakeholder needs.

Case studies of inclusive rigour in practice

We take a multiple case study approach (Marrelli, 2007) within which each evaluation experience is a unique case of inclusive rigour. We first provide an overview of the three cases, through describing the evaluand in context before sharing the evaluation design and results. All three projects were funded by Humanity United's peacebuilding portfolio that aims to transform the peacebuilding system by centering the agency and power of local peacebuilders, with the potential for more enduring, resilient peace. Consequently, all of the cases showcase applications of the framework to participatory interventions. In this context, the evaluation and learning designs were aiming to produce credible evidence as well as directly contributing to the transformative impact the interventions seek. All dimensions of inclusive rigour in each case are summarised in Table 2.

Mali Vestibule de la Paix

The *Vestibule de la Paix* project is a US\$3.5 million peacebuilding intervention operating in Mali (2018–2024) that aims to develop and test a participatory peacebuilding methodology based on systemic action research (SAR). SAR is a form of action research that uses peoples' lived experience as a starting point to uncover underlying dynamics that lead to a particular issue of concern in a system, in this case the manifestation of different forms of conflict. SAR generates participatory evidence, as well as actions in response to this causal understanding, with the aim of seeding change across the system (Burns, 2007). This project is the first large-scale application of SAR for peacebuilding.

The participatory process is implemented in three localities – the first with low levels of conflict and closer to the capital, and two with higher levels of conflict in the north. Through the SAR process, causal dynamics of conflict were identified and depicted on a large system map, presenting the main dynamics that action research groups chose to respond to. The action research groups include diverse local actors who facilitated their own research process to collect further evidence, and develop local-level actions.

A four-way partnership model (see details in Table 2) seeking to learn from the SAR methodology led to an embedded monitoring and evaluation system, guided by an overarching

Table 2. Summarised cases of inclusive rigour.

Case	Methodological bricolage	Meaningful participation and inclusion	Utilisation and impact	Enabling environment
Mali <i>Vestibules de la Paix</i>	Contribution analysis overarching design with causal theories of change explored through participatory process evaluation (case studies), outcome harvesting and causal analysis	Participants involved in their own evaluation of actions, which builds skills to serve them beyond the project. Local harvesters collected large numbers of outcome descriptions. Implementation team involved in analysis of outcomes and contribution claims.	Learning from the first community SAR process informed adaptation to second and third. Evaluation results provide the first evidence of systemic outcome potential of SAR in the context of peacebuilding, which can inform future programming.	The programme was co-developed and implemented through a four-way partnership between IMRAP (a Malian peacebuilding NGO), Interpeace (a Swiss peacebuilding INGO), Institute of Development Studies (a UK-based research institute with expertise in SAR and evaluation research), and Humanity United (the funder). The four organisations had different MEL cultures: some were more research and learning oriented, while others were less comfortable with emergent designs and causal emphasis and were more preoccupied with demonstrating added value and impact of the SAR approach. Others still associated MEL with performance auditing, leading to difficulties with obtaining key information and engaging key participants. Different levels of expertise in the specific MEL designs and approaches also created confusion around what was possible and what to expect. A shared governance approach created a space within which differences could potentially be navigated.
Colombia <i>Co-Inspira</i>	Causal hypotheses tested through participatory outcome harvesting, systematisation and causal analysis	Participants included a broader range of civil society actors who were exposed to causal hypotheses and involved in analysis leading to new understanding about how to create change through their work	Learning was fed into the second phase of the programme and use of SAR towards intended outcomes. Evidence on how to strengthen Peace Councils was shared with the departmental and municipal governments and the High Commissioner for Peace Office to support the Peace Councils. Evidence on the limitations and strengths of SAR as a peacebuilding methodology shared with the donor community.	The project benefitted from strong agreement between donor and implementer, both appreciative of complexity-aware participatory methods. The methodological bricolage was an intentional design supported by Humanity United, independently from the main funding of the project, which followed a more traditional approach. This enabled the team to compare across the two approaches.

(Continued)

Table 2. (Continued)

Case	Methodological bricolage	Meaningful participation and inclusion	Utilisation and impact	Enabling environment
Colombia Everyday Justice	Everyday Peace Indicators leading to photovoice on specific indicators to explore experiences in depth. Evaluative judgements came through EPI-applied pre and post intervention	Community participants involved in both EPIs and photovoice methods. Photovoice enabled inclusion of marginalised groups such as illiterate people and children.	Communities use the photovoice experience to strengthen their own engagement in building everyday peace. Partners are using methodological learning to inform future design.	The main implementing organisation, Everyday Peace Indicators, sought the expertise of photovoice practitioners to add to the EPI methodology used in all locations. Everyday Peace Indicators retained the lead and control for both overall design and operational decisions. A key motivation for combining the methods was to enhance participation, build community ownership and encourage action around EPI indicators (Fairey et al., 2022). Everyday Peace Indicators managed the design and bricolage of the EPI and photovoice components, thus decisions were based mostly on their methodological understanding and institutional priorities, and continues to explore the results of this pilot for bricolage application in their organisation.

Source: authors own.

Note: SAR = systemic action research; NGO = non-governmental organisation; MEL = monitoring, evaluation and learning; EPI = everyday peace indicator.

contribution analysis design to enable adaptive management and address the main evaluation question: ‘How does the SAR process contribute to the conditions for community-based peacebuilding?’

Detailed causal theories of change were co-developed and an extensive process documentation system was set up. A baseline and endline of all participants enabled tracking of changes in their attitudes and behaviours on qualitative domains related to conflict mediation and local agency, as per the theories of change. Case studies of both successful and unsuccessful action research groups produced an in-depth understanding of if and how they worked, in context and for whom. Participatory outcome harvesting implemented by a team of local harvesters then captured emergent outcomes beyond the documented SAR activities, and a causal analysis of all outcomes led to substantiation of claims of contribution to pathways towards systemic change.

At the time of writing, evaluation was complete in Kangaba, the first of three localities, where levels of conflict were lowest. The evidence suggests that the SAR approach, with its emphasis on causal analysis of system dynamics, and participatory conflict mediation processes, implemented by a research team that took time to work in contextualised ways, led to an increase in local capacity for management of (non-armed) conflict. Participants in the action research processes are experiencing high levels of respect and are showing signs of greater agency to engage in conflict mediation. The outcome harvesting has also shown that greater collaboration is leading to improved relations in some communities. The contribution evidence gathered suggests that these changes are in part the result of a highly contextualised design that intentionally included authoritative people from the communities in the process. It also shows that social norms related to women’s engagement were responded to but not overcome, leading to less meaningful involvement of women in the process despite intentional inclusion strategies.

Colombia Co-Inspira

The *Co-Inspira* project is implemented by the NGO AdaptPeacebuilding, beginning in three municipalities in Colombia in 2020. Following the signing of the Colombian Peace Accord in 2016, the country’s network of local peace councils have provided a space where peacebuilding issues could be presented and acted upon by elite and non-elite actors at multiple levels of society. The project design responds to factors that have hampered the peacebuilding efforts of the councils: lack of trust, political agendas influencing implementation and a severe lack of resources.

The project took advantage of two distinct funding streams to compare two alternative approaches to revitalising the work of the peace councils. In two municipalities, an externally designed capacity building and dialogue approach was employed, while in the third, a SAR approach, emphasising inclusive, participatory decision-making and action-based learning, was implemented. In the SAR approach, the timing, topics, participants, modalities and success measures are determined by the participants themselves, rather than according to the requirements, interests and assumptions of external actors alone. Life stories were collected among community members, then analysed to produce collective causal loop diagrams demonstrating the main challenges and opportunities to build peace in their territories. Based on the identified dynamics and associated theories of change, participants formed action research groups to respond to three prioritised peacebuilding opportunities: womens’ empowerment, youth solidarity and environmental protection.

The evaluation compared how the two approaches worked and whether discernible differences existed in how and to what extent local peace councils contributed to conditions of ‘everyday peace’. Causal pathways for how the two approaches would strengthen the peace councils were theorised based on relevant sociological and peace and conflict literature, as well as the empirical experience of several previous rounds of SAR for peacebuilding in Myanmar (Fray and Burns, 2021). The causal pathways theorised how peacebuilding methodologies influence the collective agency and relationships between people and organisations involved in peacebuilding. Baseline data related to these causal pathways were collected at the outset of the project via a survey of peace councillors in three locations. Towards the end of the project, participants reflected on the causal pathways, using an adapted outcome harvesting approach, first describing any outcomes that were occurring in relationship to these, then scoring their significance from a peacebuilding perspective, as well as describing the contribution of the project. The evaluation team then compared findings from the outcome harvesting process with baseline data from the survey, and supplementary evidence from a *sistematización* process (a Latin American process evaluation method – see Mera Rodríguez, 2019; Pérez de Maza, 2016).

The evaluation found that the peace council that employed the SAR approach initiated more local peacebuilding activities than other peace councils and were more widely known among the local community. Peace councillors in this location tended to be seen as more legitimate than in other locations. The evaluators acknowledged, however, that the 6 months of the pilot project was insufficient to reveal a large number of significant peacebuilding outcomes, or to comprehensively test the causal pathways. Additional data are being gathered in 2023 and will continue in 2024 after 2 years of implementation, which will allow for evidence with higher transferability potential.

Colombia Everyday Justice

The Everyday Justice project, initiated in 2019 by the NGO Everyday Peace Indicators is implemented in three regions of Colombia. The focus is on improving the integrated justice, reparation and non-repetition mechanisms set up to support the implementation of the 2016 Peace Accord. The lack of coherence across national, local and community priorities and agendas, means there is a lack of understanding around the everyday needs of communities. Everyday Justice responded to this by exploring how the communities’ experiences of transitional justice processes are affecting coexistence, feelings of justice and perceptions of institutional accountability (Dixon and Firchow, 2022).

The project includes use of participatory everyday peace indicators (EPIs). Using a community-engaged process, participants identify and agree on a set of everyday experiences (such as access to markets, trust in neighbours, etc.) that are meaningful to them as markers of a peaceful life (Firchow, 2018). These EPIs then form the basis of surveys that are used to monitor change and assess communities’ varying experiences of peacebuilding processes over time.

In two communities in each of the three regions of operation, Everyday Peace Indicators integrated the participatory indicator process with photovoice, a participatory action research method (Sutton-Brown, 2014; Wang and Burris, 1997). The two methods were sequenced, with EPI indicators used as launchpads for the creation of photo stories by members of the community inspired by specific indicators that resonated with them. Participants took photos and wrote accompanying narratives about the significance of the indicators, to

each individually and in groups. The resulting photo stories were refined through a collective review process, and a small selection was chosen for a public exhibit. In a final workshop, participants reflected on their experience of the photo story and exhibit processes.

Evaluative research conducted around the photovoice process in the first two communities revealed that the process supported healing, enabled intergenerational dialogue, built territorial identity and catalysed community peace actions (Fairey et al., 2022). Crucially, these community impacts emerged as the photovoice process built on existing community strengths and priorities as identified in the participatory indicator process.

Learning about inclusive rigour in practice

Table 2 presents a comparative view of the three unique cases. The emphasis of evaluation in all three cases was, from the outset, use-oriented. The evaluations aimed to generate learning about how approaches to peacebuilding focused on local agency were working in context. Across all, we see that an embedded evaluation design was intimately connected to the participatory nature of the interventions themselves, producing in-depth understandings of the processes through which outcomes related to conflict mediation and experiences of peace are generated, as well as directly feeding the actions taken. Furthermore, they all combined methods to collect data on pre-defined outcomes (such as through the baseline/endline designs in both the *Vestibule de la Paix* and the *Co-Inspira* evaluations and EPI in Everyday Justice) with methods that explored emergent causal pathways (through photovoice and outcome harvesting).

Our cross-case analysis generated reflections on the relationship between the three design and practice dimensions of the framework and how they are enmeshed in the characteristics of the broader environment that enable or hinder inclusive rigour. Balancing across the different dimensions is not always easy, trade-offs are common, and opportunities to reconcile in creative ways can emerge. We expand on two areas of learning that emerge from our experiences of navigating tensions and identifying opportunities for the practice of inclusive rigour.

Institutional arrangements to navigate multiple use values

Across all cases, as we might expect, the multi-partner set-up of the projects and their associated MEL systems played a major role in how easy or challenging it became to navigate different use values. This, in turn, influenced the methodological choices made as well as the levels and forms of participation they allowed.

This is perhaps most evident in the *Vestibule de la Paix* case, given the unique four-way partnership, which included vastly different organisations, and the donor involved as a partner in both governance and operational arrangements. Each of the partners brought a different form of expertise and with it, their unique perspective on the purpose of evaluation. For the Institute of Development Studies, the evaluation technical lead, the theory building opportunity, requiring in-depth and detailed documentation based on the causal theories of change, was particularly exciting, bringing a knowledge production emphasis. Other partners were most concerned with building strong evidence of the contribution of SAR towards specific peacebuilding effects to convince others of its added value. For the local peacebuilding partner, MEL was initially understood as an instrument to manage performance and later to demonstrate impact. Depending on the partner, the evaluation

design and specific bricolage was either experienced as too conceptual (by those more practically oriented) or too simple (by those more theoretically oriented) and was seen to be too focused either on the ‘how’ or the ‘what’.

Navigating these different views was challenging for the evaluation team. Commitment to the four-way governance model did, however, mean that the evaluation team included representatives of the four partners, allowing the respective viewpoints to be explored in a safe space. The team spent time on communicating and interpreting the technical nuance of what methods could enable what type of causal evidence and how they served their respective interests. While this supported rigour in the evolving methodological design, these dynamics meant much of the evaluation team’s energy was turned inwards, and as a result, missed opportunities to fully ground the purpose of evaluation with the participating communities. For example, participants could have also been considered as active partners in discussions about what evaluation could be used for, which might have deepened participation, or the analysis of outcomes within outcome harvesting could have been more inclusive of community perspectives.

In the *Everyday Justice* project, separate teams implemented the participatory indicators and photovoice processes, working with local community facilitators in each place. Alongside the operational separation were also different emphases and priorities, designed to complement and amplify each other. Given this separation, the evaluative component of the design was delivered through the EPI method alone, while the photovoice component focused on knowledge generation for community action. The EPI team implemented a post-project end-line and compared it with the initial EPI baseline to detect changes in everyday peace outcomes. Sometimes they used photos from the initial photovoice component to illustrate some of their indicators. But an additional photovoice workshop to update on communities’ vision of their situation post-project was not implemented. Both EPI and photovoice partners are currently in discussion on how to strengthen their methodological bricolage for evaluation to enhance causal inference through different forms of evidence. They acknowledge and are actively exploring the need to build a more enabling environment for combining different forms of evidence and appreciating quality across them.

In the case of *Co-Inspira*, common practice and shared learning agendas between partners enabled the bricolage set up to respond to the different needs. It was able to build theoretical understanding related to questions of power and agency in the causal pathways which was a particular research interest for some. Further, the partners and the donor were all keen to learn about the methodology (SAR) and to identify and explore emergent outcomes. Finally, the SAR embedded design meant that participants could also learn about different levels of the peacebuilding processes. In this case, the commitment to ongoing work with the peace councillors created the structures for the team to prioritise participant expectations in the knowledge that this would allow adaptation to subsequent phases of work.

Evaluation cultures and mindsets mediating uncertainty

Behind and within the partnership arrangements sit the cultures and mindsets of evaluation. Experiences of top-down aid are often internalised by actors throughout the peacebuilding system and are expressed through ways of organising that assume a command-and-control approach based on predefining all activities, outputs and outcomes. Methodological bricolage, on the contrary, requires openness to flexibility, as well as the ability to manage uncertainties

along outcome pathways. Our experiences show that different levels of comfort with uncertainty and flexibility underpinned partners' appetite for emergent and ongoing co-design. And this, in turn, influenced our ability to maintain quality and intentionality with methodological bricolage.

The Everyday Justice experience illustrates that when participation is meaningful, and evaluation methodologies are adapted to the local context, this contextualised view drives how methods are combined. In the communities where EPI and photovoice were implemented, on average, half of the photovoice participants had been involved in the EPI process. In these communities, locals did not engage with them as distinct methods, but rather, they were understood as one joined up process driven by their lived experiences in spite of the separation in operational terms. A holistic local view of what matters was driving methodological decisions *in situ*. It also, however, requires a high degree of openness to uncertainty, as it is not possible to know exactly what will emerge and where participants will drive the process.

One of photovoice's main qualities is the physical output created by the communities themselves, using photographic exhibitions that symbolise different perceptions of justice and conflict. This open participatory aspect comes with a certain level of security risk for those sharing photographs. In some cases, photos had to be presented in safe spaces instead of on the walls of schools, community centres or other buildings as was originally planned. In other cases, certain photos could not be put up at all in order to ensure protection of participants (Fairey et al., 2022). In the context of peacebuilding, navigating emergent risk through a participatory process is a key skill set that embedded evaluators need.

Comfort levels with emergent design were not always as high as required for smooth bricolage practice. In all three cases, the evaluand itself is defined by participants and so cannot be known fully *de facto*. In the *Vestibule de la Paix* case, some partners assumed a linear approach to management through predefined and planned activities, outputs and expected outcomes. Lack of evaluation tools they were familiar with, such as logframes and clear SMART predefined indicators, led to discomfort by management in some instances. In response to this pressure, the evaluation team produced a design document with a best guess of all methods and outputs that would be required, even as causal theories of change were evolving on the ground. The fact that methodological bricolage was enshrined in a formal document provided the evaluation team the means through which to share the logic of combining methods within an overarching Contribution Analysis design. This explicit and intentional view of the design reassured some partners and provided a robust alternative to what they were expecting.

Yet for the evaluation team, this initial high-level design was less useful when iterative recombining and adaptation of methods was required to meet shifting demands and contextual conditions. Approaches and choices were being revised constantly based on information coming from the field and on the partners general view of what the project sought to achieve in context. In practice, for the evaluation team, making the design explicit (and getting it agreed) required a considerable up-front investment in time (when specifics were impossible to fully know) for relatively low return, and served more of a pacifying function than a technical one. The evaluation mindset of some partners had to be overcome in order to get to the work of actually ensuring rigour in the design.

In the experience of *Co-Inspira*, both the donor and the implementing organisation were appreciative of and wanting to experiment with complexity-aware participatory methods. Yet external partners, especially the local authorities were hesitant at first to engage with the team

and discuss the effects of the SAR methodology. Difficulties arose as it became clear that some partners were listening for specific results that they wanted, rather than interested to find out more about the changes that did, in fact, occur. The team found that this dynamic shifted somewhat once the stories of change (from outcome harvesting) could be fully told and through them contributions could be explained via the richness of local explanations and documentation sources provided by the team. Across all cases, we see that narratives of change are powerful tools as part of the bricolage design.

Conclusion

While the field of peacebuilding is intentionally moving towards locally led, participatory and adaptive programming, the evaluation of peacebuilding interventions is lagging behind. It has largely remained trapped by narrow definitions of rigour, stemming from unhelpful hierarchies of knowledge that lead to a focus on measuring predefined indicators rather than building causal explanations. The inclusive rigour framework responds to this challenge. It builds on existing trends in evaluation, that argue for complexity-aware and epistemologically plural approaches as appropriate, to build credible causal explanations in conditions of uncertainty. It identifies three domains of design and practice that need to be considered together to build rigour throughout ongoing evaluation design and in its implementation. Rigour here is not defined by methodological choice alone, but rather, relies on an active view of evolving methodological choices that unfold throughout an iterative process as maximum use value and meaningful participation are sought.

As we applied the framework to our own work, we learned how critical the institutional arrangements are in creating the enabling environment that sits behind and helps to work across the three domains. We have shown how the partnership dynamics and decision-making mechanisms in our cases drove whether it was possible to balance different needs, spanning accountability, learning, action and evidence and knowledge production, through using multiple methods and paying attention to inclusion, and where trade-offs could not be navigated successfully. We also show where evaluation cultures and mindsets were not aligned with a more inclusive view of rigour, the depth and form of participation were limited as well as the ability to practise methodological bricolage, and so to sustain credibility in building causal explanations of emergent outcome pathways.

Underlying both of these dimensions is the implicit, and at times explicit, presence of power dynamics. As has been described by others (Baur et al., 2010; Roche and Kelly, 2012) underpinning the formal institutional arrangements (structures) and mostly informal (and often hidden) values, cultures and mindsets of partners involved in an evaluation process, lie different forms and levels of power that mediate how decisions about use, participation and methods are made. Power asymmetries might exist along a number of lines, some embedded in historical colonial legacies (perpetuated by top-down hierarchies of aid), others related to different social norms embedded in context (related to gender or class for example) and yet others link to valuing different forms of knowledge (such as valuing experimental designs over other causal designs). These power dynamics of development and evaluation practice are also, in fact, part of the broader systems of governance that are the focus of peacebuilding interventions. Yet the way in which they influence both achievement of peacebuilding outcomes, and achievement of quality in evaluation practice, are often overlooked in design and operationalisation. In the realm of evaluation, an over-emphasis on the technical, and the

focus of rigour as linked entirely to initial methodological choice, tends to overshadow these more hidden dynamics.

We reflected in our cases on how navigating power is central to an inclusive rigour practice although we did not agree to use of one particular theory of power. In both the Everyday Justice and *Vestibule de la Paix* cases, there was not sufficient attention paid to power within the partnership, and hidden dynamics were not able to be brought out into the open to be discussed and potentially navigated. In Mali for example, there was an attempt to use a partnership rubric to build reflexivity on the partnership dynamic itself, but was not prioritised among competing demands on time. It is likely that partners had no appetite for uncomfortable conversations.

In our third case, the *Co-Inspira* project, however, the causal theories of change that were being tested through the evaluation included theories about how power sharing occurs in the context of the Peace Councils. Making these theories explicit to participants during the evaluation process, as a way to democratise evaluation, led to deeper insights across all partners about how power was also showing up in their own practice. Participants reflected on this being a key 'aha' moment for them in their own internalisation and reflection on their power to create change, and expanded their understanding of what the evaluation was trying to capture: the many changes beyond the simple predefined 'outputs' and 'indicators' they were used to measuring.

This experience illustrates the potential of being intentional in creating a safe-enough container for hidden dynamics to be surfaced and discussed. This is one way to build reflexivity, which is a core skill set for acknowledging and then finding ways to navigate power dynamics that might otherwise derail attempts to practise inclusive rigour. It also illustrates what we intend the framework to do – to inspire new lines of questioning as peacebuilding evaluators and implementers build intentionality in moving towards more helpful framings of rigour, and ultimately, use evaluation to build credible causal explanations in conditions of uncertainty.

From the learning presented in this article, several lines of new and ongoing inquiry can focus new rounds of action-oriented learning about how to shift the way we conceptualise and operationalise rigour in peacebuilding evaluation: (1) how does an emphasis on credibility of appropriate causal explanations as central to methodological choice, influence the quality of participation and the potential for greater utilisation?; (2) how can we intentionally build greater openness to complexity and emergence within evaluation cultures and partnerships to enable inclusive rigour?; and related to the previous two (3) which theoretical orientations of power (of the many possible) are helpful vehicles for operationalising real-world navigation of power dynamics in evaluation?

Acknowledgements

We are grateful for the time and insights offered by the individuals, many of whom are victims of conflict, that have engaged in the participatory peacebuilding programmes and evaluation processes that have enabled the development of case studies in the application of the inclusive rigour framework.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Humanity United has funded empirical work in the three case studies.

ORCID iD

Marina Apgar  <https://orcid.org/0000-0002-5775-6007>

References

- Andersen OW, Bull B and Kennedy-Chouane M (eds) (2013) *Evaluation Methodologies for Aid in Conflict*. Routledge.
- Apgar M and Allen W (2021) Participatory monitoring evaluation and learning: Taking stock and breaking new ground. In: Burns D, Howard J and Ospina SM (eds) *SAGE Handbook of Participatory Research and Inquiry*. Thousand Oaks, CA: Sage, 829–935.
- Aston T and Apgar M (2022) The art and craft of bricolage in evaluation. Centre for Development Impact Practice Paper. Available at: <https://www.ids.ac.uk/publications/the-art-and-craft-of-bricolage-in-evaluation/>
- Aston T, Roche C, Schaaf M, et al. (2022) Monitoring and evaluation for thinking and working politically. *Evaluation* 28(1): 36–57.
- Baker S and Chapin FSI (2018) Going beyond ‘it depends’: The role of context in shaping participation in natural resource management. *Ecology and Society* 23(1): 20.
- Bamberger M, Rao V and Woolcock M (2010) Using mixed methods in monitoring and evaluation: Experiences from international development. World Bank Development Research Group Poverty and Inequality Team Policy Research Working Paper. Available at: <https://openknowledge.worldbank.org/entities/publication/0ccee604-f1c5-5c76-9852-3f3d9f7f5d7e>
- Baur VE, Van Elteren AHG, Nierse CJ, et al. (2010) Dealing with distrust and power dynamics: Asymmetric relations among stakeholders in responsive evaluation. *Evaluation* 16(3): 233–48.
- Befani B (2020) *Choosing Appropriate Evaluation Methods – A Tool for Assessment and Selection* (Version Two), Guildford: Centre for the Evaluation of Complexity Across the Nexus. Available at: https://www.cecan.ac.uk/wp-content/uploads/2020/11/Final_Choosing-Appropriate-Evaluation-Methods-1.pdf
- Befani B, Ramalingam B and Stern E (2015) Introduction – Towards systemic approaches to evaluation and impact. *IDS Bulletin* 46(1): 1–6.
- Bornstein L (2010) Peace and Conflict Impact Assessment (PCIA) in community development: A case study from Mozambique. *Evaluation* 16(2): 165–76.
- Brisolara S, Seigart D and SenGupta S (2014) *Feminist Evaluation and Research: Theory and Practice*. New York: Guilford Publications.
- Burns D (2007) *Systemic Action Research: A Strategy for Whole System Change*. Bristol: Bristol University Press.
- Campbell DT and Stanley JC (1963) Experimental and quasi-experimental designs for research. In: Bell C, Gitomer D and Gitomer DH (eds) *Handbook of Research on Teaching*. Chicago, IL: Rand McNally College Publishing, 171–246.
- Chambers R (2015) Inclusive rigour for complexity. *Journal of Development Effectiveness* 7(3): 327–35.
- Chilisa B (2019) *Indigenous Research Methodologies*. Thousand Oaks, CA: Sage.
- Chilisa B and Mertens DM (2021) Indigenous made in Africa evaluation frameworks: Addressing epistemic violence and contributing to social transformation. *American Journal of Evaluation* 42(2): 241–53.
- Chouinard JA (2013) The case for participatory evaluation in an era of accountability. *American Journal of Evaluation* 34(2): 237–53.
- Cousins JB and Whitmore E (1998) Framing participatory evaluation. *New Directions for Evaluation* 80: 5–23.
- Davidson EJ (2013) *Actionable Evaluation Basics: Getting Succinct Answers to the Most Important Questions*. Seattle, WA: Real Evaluation Limited.
- Delgado C, Milante G, Riquier M, et al. (2022) *Measuring Peace Impact: Challenges and Solutions*. Stockholm: SIPRI.
- Dixon P and Firchow P (2022) Collective justice: Ex-combatants and community reparations in Colombia. *Journal of Human Rights Practice* 14(2): 434–453.

- Downes J and Gullickson AM (2022) What does it mean for an evaluation to be ‘valid’? A critical synthesis of evaluation literature. *Evaluation and Program Planning* 91: 102056.
- Eager R and Barnett C (2021) Evidencing the impact of complex interventions: The ethics of achieving transformational change. In: Stame N, Hawkins P and Van den Berg RD (eds) *Ethics for Evaluation* (pp. 124–140). London: Routledge.
- Eyben R, Gujit I, Roche C, et al. (2015) *The Politics of Evidence and Results in International Development*. Rugby: Practical Action Publishing.
- Fairey T and Kerr R (2020) What works? Creative approaches to transitional justice Bosnia and Herzegovina. *International Journal of Transitional Justice* 14(1): 142–64.
- Fairey T, Firchow P and Dixon P (2022) Images and indicators: Mixing participatory methods to build inclusive rigour. *Action Research*. Epub ahead of print 11 October. DOI: 10.1177/14767503221137851.
- Fetterman DM (1994) Empowerment evaluation. *Evaluation Practice* 15(1): 1–15.
- Firchow P (2018) *Reclaiming Everyday Peace: Local Voices in Measurement and Evaluation After War*. Cambridge: Cambridge University Press.
- Firchow P and Selim Y (2022) Meaningful engagement from the bottom-up? Taking stock of participation in transitional justice processes. *International Journal of Transitional Justice* 16(2): 187–203.
- Forsyth J, McKee T and Benson A (2021) Data, Development discourse, and decolonization: Developing an indigenous evaluation model for indigenous youth Hockey in Canada. *Canadian Ethnic Studies* 53(3): 121–40.
- Fransman J, Hall B, Hayman R, et al. (2021) Beyond partnerships: Embracing complexity to understand and improve research collaboration for global development. *Canadian Journal of Development Studies / revue canadienne d'études du développement* 42(3): 326–46.
- Friedman VJ (2008) Action science: Creating communities of inquiry in communities of practice. In: Bradbury H and Reason P (eds) *Handbook of Action Research: The Concise Paperback Edition*. London: Sage, 131–43.
- Gates E and Dyson L (2017) Implications of the changing conversation about causality for evaluators. *American Journal of Evaluation* 38(1): 29–46.
- Gibson CMG (2018) *Deciding Together: Shifting Power and Resources Through Participatory Grantmaking*. New York: GrantCraft.
- Gleditsch NP, Nordkvelle J and Strand H (2014) Peace research – Just the study of war? *Journal of Peace Research* 51(2): 145–58.
- Guba EG and Lincoln YS (1989) *Fourth Generation Evaluation*. Thousand Oaks, CA: SAGE.
- Hanberger A (2022) Power in and of evaluation: A framework of analysis. *Evaluation* 28(3): 265–83.
- Hargreaves M (2021) Bricolage: A pluralistic approach to evaluating human ecosystem initiatives. *New Directions for Evaluation* 2021(170): 113–24.
- HM Treasury (2020) *Magenta Book 2020: Supplementary Guide: Handling Complexity in Policy Evaluation*. London: HM Treasury.
- House ER (2014) Origins of the ideas in *evaluating with validity*. *New Directions for Evaluation* 2014(142): 9–15.
- Hurteau M and Williams DD (2014) Credible judgment: Combining truth, beauty, and justice. *New Directions for Evaluation* 2014(142): 45–56.
- Jenal M and Liesner M (2017) Causality and attribution in market systems development. BEAM Exchange. Available at: <https://beamexchange.org/resources/950/>
- Jiménez-Buedo M and Russo F (2021) Experimental practices and objectivity in the social sciences: Re-embedding construct validity in the internal–external validity distinction. *Synthese* 199(3–4): 9549–79.
- Johnson K, Greenesid LO, Toal SA, et al. (2009) Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation* 30(3): 377–410.

- Kelly LM and Htwe PPT (2023) Decolonizing community development evaluation in Rakhine state, Myanmar. *American Journal of Evaluation*. Epub ahead of print 13 February. DOI: 10.1177/10982140221146140.
- Lévi-Strauss C (1966) *The Savage Mind*. Chicago, IL: University of Chicago Press.
- Lynn J and Apgar M (forthcoming) Exploring causal pathways amid complexity understanding when and how causality can be made visible. In: Newcomer K and Mumford S (eds) *Research Handbook on Programme Evaluation*. Cheltenham: Edgar Elgar Publishing.
- Marrelli AF (2007) Collecting data through case studies. *Performance Improvement* 46(7): 39–44.
- Maxwell JA (2004) Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher* 33(2): 3–11.
- Mera Rodríguez A (2019) La sistematización de experiencias como método de investigación para la producción del conocimiento. *ReHuSo: Revista de Ciencias Humanísticas y Sociales* 4(1): 113–23.
- Miller C and Haylock L (2014) Capturing changes in women’s lives: The experiences of Oxfam Canada in applying feminist evaluation principles to monitoring and evaluation practice. *Gender and Development* 22(2): 291–310.
- National Institute for Health and Care Excellence (NICE) (2012) *Methods for the Development of NICE Public Health Guidance*. London: NICE.
- Oakley AA (2022) “Politics is more difficult than physics”: Complexity and the challenge of democracy, human rights, and governance program evaluation. *New Directions for Evaluation* 176: 15–32.
- Paffenholz T (2015) Unpacking the local turn in peacebuilding: A critical assessment towards an agenda for future research. *Third World Quarterly* 36(5): 857–74.
- Parkhurst J (2017) *The Politics of Evidence: From Evidence-Based Policy to the Good Governance of Evidence*. Taylor & Francis.
- Pasanen T and Barnett I (2019) Supporting adaptive management. ODI Working Papers. Available at: <https://odi.org/en/publications/supporting-adaptive-management-monitoring-and-evaluation-tools-and-approaches/>
- Patton MQ (2002) Feminist, yes, but is it evaluation? *New Directions for Evaluation* 2002(96): 97–108.
- Patton MQ (2019) *Blue Marble Evaluation: Premises and Principles*. New York: Guilford Publications.
- Pattyn V and Bouterse M (2020) Explaining use and non-use of policy evaluations in a mature evaluation setting. *Humanities and Social Sciences Communications* 7: 1–9.
- Pearson d’Estree T (ed.) (2020) *New Directions in Peacebuilding Evaluation*. London: Rowman & Littlefield.
- Pérez de Maza T (2016) *Guía Didáctica para la Sistematización de Experiencias en Contextos Universitarios*. Caracas, Venezuela: Universidad Nacional Abierta.
- Podems DR (2010) Feminist evaluation and gender approaches: There’s a difference? *Journal of Multidisciplinary Evaluation* 6(14): 141–17.
- Preskill H and Lynn J (2016) Redefining rigour: Describing quality evaluation in complex adaptive settings. FSG Reimagining Social Change. Available at: <https://www.fsg.org/blog/redefining-rigor-describing-quality-evaluation-complex-adaptive-settings/>
- Price R, Snijder M and Apgar M (2021) *Defining and evaluating equitable partnerships: A rapid review*. Working Paper, March. Tomorrow’s Cities project team, Nairobi, Kenya.
- Prieto-Martín P, Faith B, Hernandez K, et al. (2017) *Doing Digital Development Differently: Lessons in Adaptive Management From Technology for Governance Initiatives in Kenya*. Brighton: Institute of Development Studies.
- Ramalingam B, Barnett I, Levy A, et al. (2017) Bridging real-time data and adaptive management: Ten lessons for policy makers and practitioners. Available at: [usaid.gov/digital-development/rtd4am/policy-design-lessons#:~:text=Information%20is%20organized%20into%20ten,different%20levels%20of%20decision%20making](https://www.usaid.gov/digital-development/rtd4am/policy-design-lessons#:~:text=Information%20is%20organized%20into%20ten,different%20levels%20of%20decision%20making).
- Ramalingam B, Wild L and Buffardi AL (2019) Making adaptive rigour work: Principles and practices for strengthening monitoring, evaluation and learning for adaptive management. ODI Briefing

Note. Available at: https://www.google.com/search?q=Making+adaptive+rigour+work%3A+Principles+and+practices+for+strengthening+monitoring%2C+evaluation+and+learning+for+adaptive+management.+ODI+Briefing+Note&rlz=1C1GCEU_enIN853IN853&oq=Making+adaptive+rigour+work%3A+Principles+and+practices+for+strengthening+monitoring%2C+evaluation+and+learning+for+adaptive+management.+ODI+Briefing+Note&aqs=chrome..69i57.731j0j4&sourceid=chrome&ie=UTF-8

- Ribeiro G (2019) Relevance, probative value, and explanatory considerations. *The International Journal of Evidence & Proof* 23(1–2): 107–13.
- Roche C and Kelly L (2012) The evaluation of politics and the politics of evaluation. Developmental Leadership Programme Background Paper 11. Available at: <https://dlprog.org/publications/background-papers/the-evaluation-of-politics-and-the-politics-of-evaluation/>
- Scharbatke-Church C (2011) Evaluating peacebuilding: Not yet all it could be. In: Austin B, Fischer M and Giessmann HJ (eds) *Berghof handbook for conflict transformation*. Berlin: Berghof Research Centre for Constructive Conflict Management, 460–480. Available at: https://berghof-foundation.org/files/publications/scharbatke_church_handbook.pdf
- Schwandt TA and Gates EF (2021) *Evaluating and Valuing in Social Research*. New York: Guilford Publications.
- Scriven M (1995) The logic of evaluation and evaluation practice. *New Directions for Evaluation* 1995(68): 49–70.
- Snijder M, Steege R, Callander M, et al. (2023) How are research for development programmes implementing and evaluating equitable partnerships to address power asymmetries? *The European Journal of Development Research* 35(2): 351–79.
- Stern E, Stame N, Mayne J, et al. (2012) *Broadening the Range of Designs and Methods for Impact Evaluations*. Brighton: Institute for Development Studies.
- Sutton-Brown CA (2014) Photovoice: A methodological guide. *Photography & Culture* 7(2): 169–85.
- Ton G (2012) The mixing of methods: A three-step process for improving rigour in impact evaluations. *Evaluation* 18(1): 5–25.
- Trust Based Philanthropy Project (2023) Trust-based philanthropy: An approach. Available at: <https://www.trustbasedphilanthropy.org/>
- Tuck E and Yang KW (2012) Decolonization is not a metaphor. *Decolonization: Indigeneity, Education & Society* 1(1): 1–40.
- Urwin E, Botoeva A, Arias R, et al. (2023) Flipping the power dynamics in measurement and evaluation: International aid and the potential for a grounded accountability model. *Harvard Negotiation Journal Special Issue: Localization and the Aid Industry* 39(4): 401–26.
- Van Hemelrijck A and Guijt I (2016) Balancing inclusiveness, rigour and feasibility: Insights from participatory impact evaluations in Ghana and Vietnam. Centre for Development Impact Practice Paper 14. Available at: <https://www.ids.ac.uk/publications/balancing-inclusiveness-rigour-and-feasibility-insights-from-participatory-impact-evaluations-in-ghana-and-vietnam/>
- Wang C and Burrell MA (1997) Photovoice: Concept, methodology, and use for participatory needs assessment. *Health Education & Behavior* 24(3): 369–87.
- White H (2019) The twenty-first century experimenting society: The four waves of the evidence revolution. *Palgrave Communications* 5(1): 1–7.

Marina Apgar is a human ecologist with interest in complexity science and inequality and conducts evaluation research on complexity-aware and participatory evaluation methods as a research fellow at the Institute of Development Studies.

Helene Bradburn is an independent MEL professional focused on evaluation in the context of peacebuilding interventions.

Livia Rohrbach is a political scientist with specialisation in comparative politics and conflict studies and is currently Search for Common Ground's regional research and evaluation coordinator in the Horn of Africa.

Leslie Wingender is a director at Humanity United where she leads peacebuilding programming and has a history of working in evaluation in humanitarian and development NGOs.

Edwin Cubillos Rodriguez is a Colombian photovoice practitioner working across a range of peacebuilding and development programmes.

Angela Maria Baez Silva Arias has a degree in cultural biology, is an expert in complexity and development leads work on systemic action research for Adapt Peacebuilding in Colombia.

Alamoussa Dioma is the MEL lead for the Vesitbule de la Paix project using systemic action research in Mali.

Tiffany Fairey is a visual sociologist working on peacebuilding, conflict transformation, reconciliation and the relationship between photography and peace.

Stephen Gray is a peacebuilding scholar, currently undertaking a PhD in local agency and power in relationship to peace efforts and is the director of Adapt Peacebuilding.

Ayak Chol Deng Alak is a peacebuilding leader, and focuses on supporting enabling environments for negotiations, mediation and dialogues in South Sudan.

Steff Deprez is an independent evaluator focusing on creative use of narratives in participatory MEL practice, and is a leading SenseMaker practitioner in international development.