

# How do we define and support quality and rigor in Causal Pathways evaluation?

*January 2025*

Marina Apgar and Tom Aston

with Jewlya Lynn, Carolina De La Rosa Mateo, Carlisle Levine, Julia Coffman, Abdoul Karim Coulibaly, Yulianto Dewata, Heather Britt



## Contents

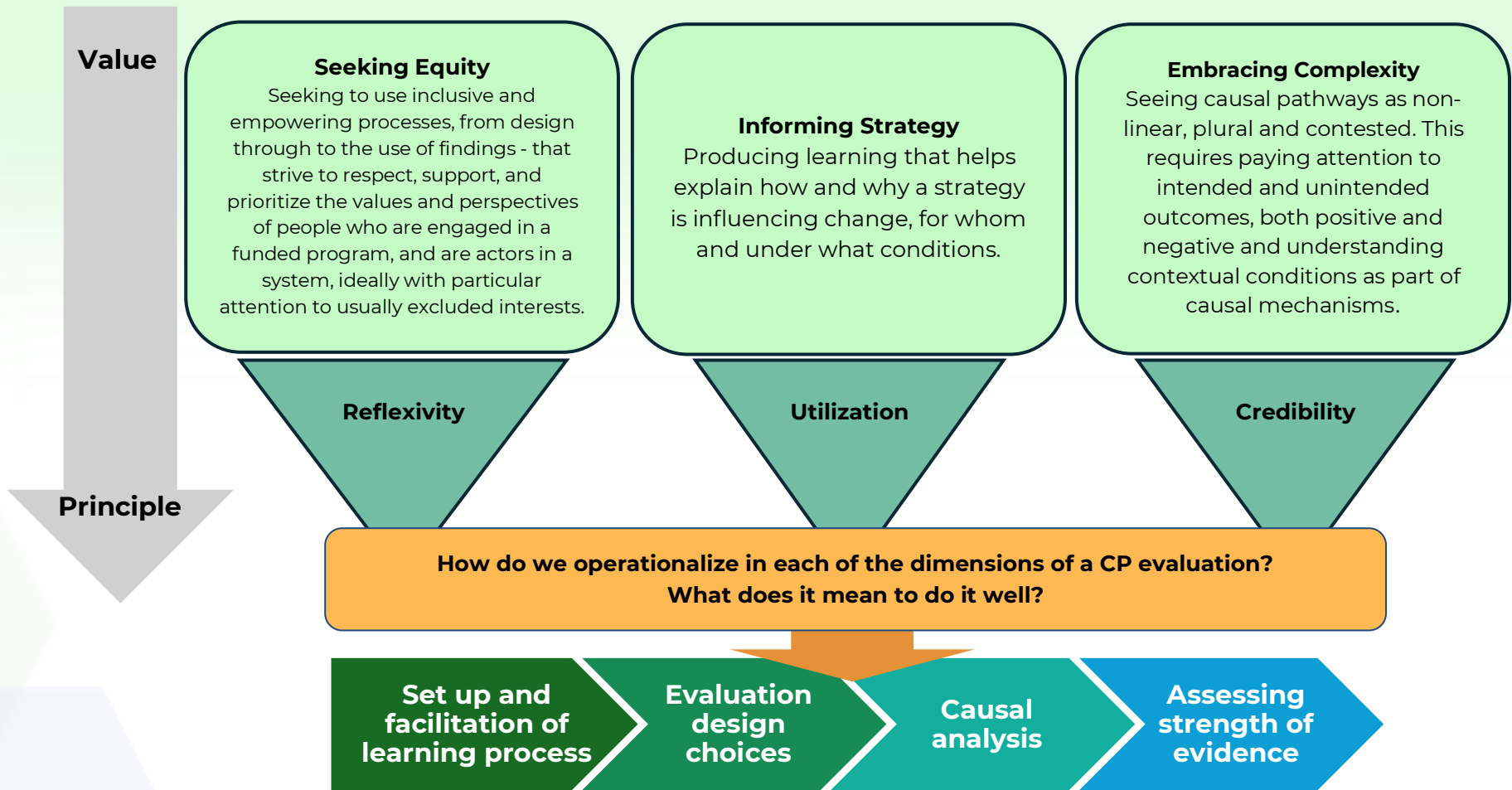
<b>About this guidance</b>	<b>2</b>
<b>What values underpin a causal pathways approach?</b>	<b>4</b>
<b>What principles help us translate values into practice?</b>	<b>5</b>
<b>How do we attend to quality through stages of evaluation?</b>	<b>7</b>
<b>Stage 1: Setting up the learning process and creating the conditions for quality</b>	<b>7</b>
<b>Stage 2. Making evaluation design choices</b>	<b>17</b>
<b>Stage 3. Implementing causal analysis</b>	<b>20</b>
<b>Stage 4. Assessing the strength of evidence of causal claims</b>	<b>24</b>
<b>References</b>	<b>25</b>

## About this guidance

**This guidance** is aimed at evaluation and learning officers in philanthropic foundations who are already using or are interested in using causal pathways as an approach to evaluation. The aim is to guide users on what 'good enough' looks like when they are: (i) using an iterative design to causal pathways evaluation; and (ii) applying causal methods to explore causal pathways in evaluating philanthropic programs and strategy.

The guidance takes the approach presented in [an article](#) published in a recent issue of the Foundation Review on Democracy, Equity and Power. It argues for taking a step back from starting with established evaluation standards and rigor criteria. We propose that it is necessary to first **define what we value when taking a causal pathways approach** to evaluation. Based on defining our values, we can then articulate **principles in alignment with the stated values** - principles explain what we mean by quality and rigor - to then be operationalised through stages of the evaluation process. Our definition of quality and rigor for causal pathways evaluation, therefore, is the result of this approach.

**Figure 1. Values defining principles to be operationalised through this guidance**



**We assume that** causal pathways evaluations are always designed in response to specific conditions including the organisational and team dynamics and experiences, the strategic learning opportunity and the type of program. Accordingly, this guidance does not present a checklist or simple recipe for how 'good enough' is achieved, but, rather **a roadmap for how to attend to quality throughout stages of the evaluation and at key decision points within them.** Ideally, teams who are new to causal pathways evaluation, should first consult the [Resource Hub on Better Evaluation](#) which includes links to additional resources. This guidance is intended to be used **as part of a facilitated process** working in collaboration with team members who have experience in choosing and using causal methodologies.

## What values underpin a causal pathways approach?

### Seeking greater equity

Causal pathways evaluations seek to increase equity in the evaluation process and where possible to turn evaluation into an empowering process. Evaluators and those involved strive to respect, support, and prioritize the values and perspectives of all people who are engaged in the funded program or strategy being evaluated. This can include participants, on-the-ground program staff, implementing partners, and others contributing to or impacted by changes in the system. This requires paying particular attention to usually excluded interests, recognising that inequities experienced in systems today are the result of coloniality and other forms of historical and ongoing systems of oppression. Defining who the main actors and owners of an evaluation are and exploring how power influences their ability to engage meaningfully is fundamental to design evaluation and learning processes that push for greater equity.

### Informing Strategy

A causal pathways approach to evaluation is well suited to answering causal questions about if, how, why, for whom and under what conditions philanthropic strategy is contributing to outcomes, and, how, in some cases, these outcomes relate to achieving systems change. These questions aim to move evaluation of strategy beyond simply describing outcomes, to exploring causality and the role of programs and activities within a mix of contextual conditions. Understanding causal relationships is fundamental to being able to make a causal claim to then

support actionable recommendations. Such exploration can enable program teams to better understand not just what their strategy is achieving, but how they can design strategy in ways that work with both opportunities and barriers to change. All causal pathways evaluations, therefore, value learning about causality as critical to building more effective strategies that contribute to systems change.

## Embracing complexity

A causal pathways evaluation begins from the premise that understanding how outcomes emerge and for whom requires first acknowledging that change is not necessarily linear - there are always multiple interacting agents that together influence complex processes of change. Causal pathways are long and can be unpredictable, influenced by non-linear system dynamics. Further, there may be multiple causal pathways that influence each other and different stakeholders might hold different views of how change happens or on what change matters and why. Embracing this complexity calls for evaluation approaches that: pay attention to a range of outcomes and impacts, both positive and negative, intended and unintended; identify and describe differences in how interventions work in different contexts for different people; pay attention to power and whose experiences are centered.

## What principles help us translate values into practice?

We define three principles that when operationalised through the stages of a causal pathways evaluation can help to build a practice which is aligned with the core values of **seeking greater equity**, **informing strategy** and **embracing complexity**. We recognise that both conceptually and practically the principles are interconnected and build upon each other.

**Reflexivity for causal pathways evaluation:** the individual and collective practices to recognise positionality and bias, and to self-consciously critique how subjectivity and context influence the evaluation process.

**Reflexivity** is a concept and practice that comes from qualitative research where it is defined as “a set of continuous, collaborative, and multifaceted practices through which researchers self-consciously critique, appraise, and evaluate how their

subjectivity and context influence the research processes.” ([Olmos-Vega et al. 2022](#)). Building awareness of self as an evaluator (knowledge, skills, and dispositions), and reflection on personal evaluation practice (competencies and areas for growth) (e.g. van Draanen, 2017) are recognised as important, yet this falls short of a deeper reflection on positionality and bias. Striving for equity requires that evaluators actively reflect on their own biases, build awareness of their positionality and power within the systems they operate.

**Utilization for causal pathways evaluations:** ensuring that evaluation findings are useful to an expanded group of ‘users’ through considering all potential system stakeholders, and striving to share power at the decision making table with program teams and participants.

**Utilization** is a common principle made popular by Michael Quinn Patton through his promotion of [utilization-focused evaluation](#) which he defines as *judging an evaluation based on its usefulness to its intended users*. In the context of causal pathways evaluation this requires not simply defining the intended users and understanding their needs, but also engaging them throughout the process. Valuing equity, and striving to not just acknowledge but also to address power relations, however, requires balancing the instrumental ‘use’ orientation, that tends to place the emphasis on the commissioner of an evaluation, which may perpetuate existing power dynamics, to recognise how power influences even who is defined as a ‘user’. An equity orientation to ‘use’, therefore, might call for usually excluded populations to be sitting at the ‘user’ decision making table. For causal pathways we define utilization in this expanded way.

**Credibility for causal pathways evaluation:** trustworthiness of the evaluation findings achieved through evaluation designs and practices that support quality causal analysis.

**Credibility** is a commonly agreed quality criterion for evaluation, which generally refers to “the trustworthiness of the evaluation findings, achieved through high-quality evaluation processes, especially rigor, integrity, competence, inclusion of diverse perspectives, and stakeholder engagement” ([BetterEvaluation](#)). What

builds credibility, therefore, is the result of a number of processes and practices. In conditions of complexity, we contend that credibility should encompass a deeper exploration of causality. In the context of causal pathways evaluation, therefore, credibility is the result of methodological designs and practices that support quality causal analysis in order to move beyond simply describing outcomes to examine how change happens, inclusive of intentional interventions and contextual factors. Causal pathways evaluation designs and methods can draw upon multiple strategies for causal analysis, requiring intentional iterative design choices throughout the evaluation.

## How do we attend to quality through stages of evaluation?

To operationalize the three principles we provide a set of operational criteria and illustrate how these can be applied to four stages of a causal pathways evaluation.



We present these stages sequentially for the sake of simplicity and clarity, however, we recognise that in reality there are often iterations and feedback loops between them.

### Stage 1. Setting up the learning process and creating the conditions for quality

**The purpose** of this initial stage, as with all evaluations, is to create the conditions that will support a high quality design and ultimately, an effective evaluation process. From the outset of a causal pathways evaluation, those involved should be intentional about **utilization** and **reflexivity** **shaping the design of the learning process** and how it will be facilitated.

↳ *Who are all the relevant actors?*

Before initiating a causal pathways design process it is critical to first consider who are all the relevant actors, paying particular attention to those often excluded. Several approaches to identifying and mapping evaluation stakeholders could be used in this phase, such as:

- The **power and interest grid** is a way to analyse stakeholders and determine how to engage them based on the levels of influence and



interest each has on the causal pathways evaluation process. See a [short video on youtube](#) introducing the tool or more on using alignment and interest grids in development of theories of change by [Clark and Apgar](#) (2019).

- The **participatory network mapping** approach - see [netmap toolkit](#) creates influence network maps to visualise how different actors influence specific outcomes (Schiffer & Hauck, 2010).

## *Case Example*

### **Realist Evaluation of The Code of Conduct for Protection of Children from Sexual Exploitation in Travel and Tourism**

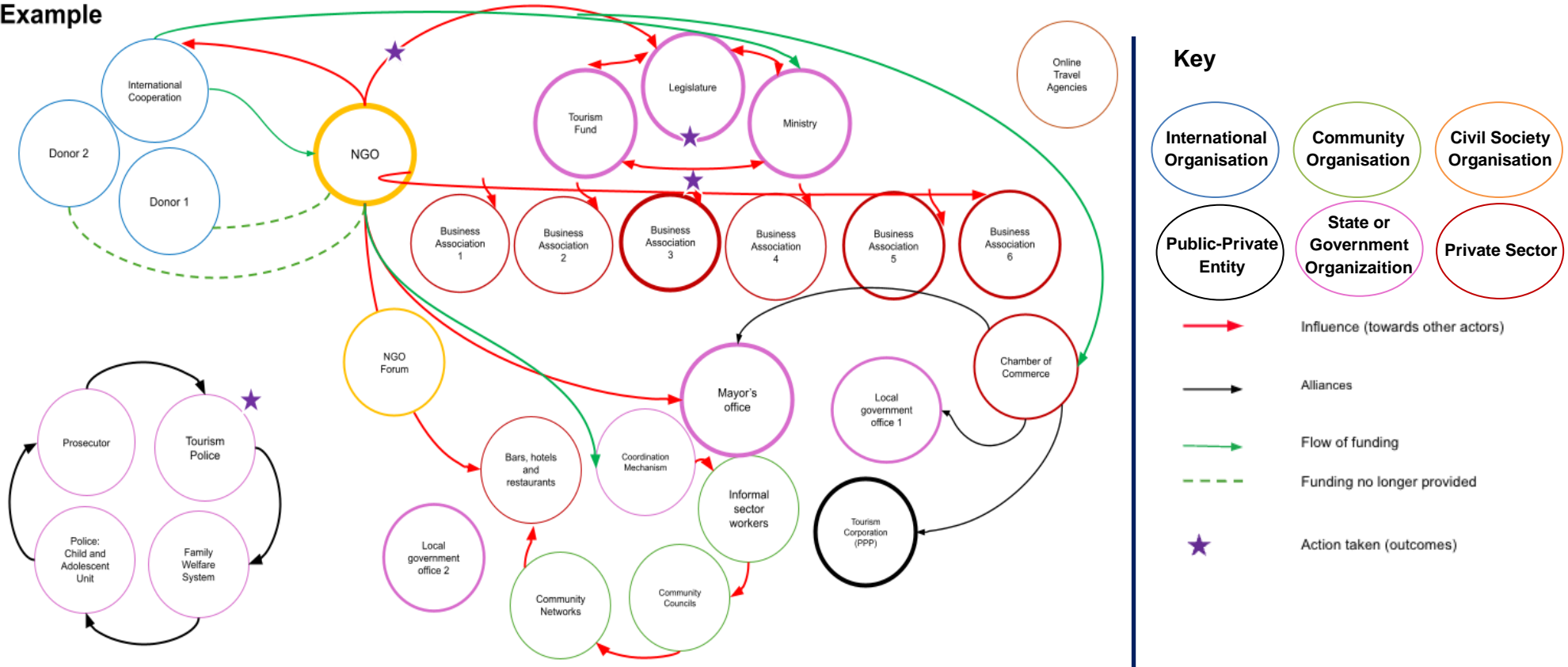
*by Alex Shirley*

In an outcome evaluation of [The Code of Conduct for the Protection of Children from Sexual Exploitation in Travel and Tourism](#), an adaptation of the netmapping tool was used in the early phases to map existing relationships between the actors that are part of a complex ecosystem working towards preventing sexual exploitation of children in the focal cities. The evaluation was commissioned by ECPAT international who coordinates implementation of The Code. The evaluation team was composed of an internal MEL specialist, an external independent evaluator and an external methodological advisor. The commissioning team within ECPAT, together with the evaluation team discussed who needed to be involved in the evaluation that aimed to produce learning to inform the strategic direction of The Code, and to directly feed into improving the design and delivery of activities to promote The Code. The Local Code Representatives (LCR), NGOs who support the implementation of The Code by businesses, were involved as participants in the netmapping exercise, given their own knowledge of the many state, city and business actors involved in the ecosystem.

Through a three hour workshop, the lead evaluator worked with members of the LCR to first define and then map all actors who were relevant to building a causal understanding of if and how The Code implementation is influencing achievement of intended and unintended outcomes. LCR participants mapped and defined relationships between actors (e.g. funding or technical support), categorised actors according to their levels of influence and then reflected on the ways in which outcomes had come about as a result of these relationships.

As shown in the map below, the thickness of circles correspond with the influence rating as assessed by the LCR (on a scale of 1 to 4, with 4 being the highest rating). During the rating exercise the LCR prioritized organisations who they felt exerted some kind of influence on others. All other actors are considered to have a rating of less than 1. Influence was defined as either the power to enforce the law, resources that an organisation could provide, or the power to influence the actions of others (e.g. hotel associations requiring members to join The Code). Influence was also defined in terms of scale and scope, where the central government was seen to exert influence over a wide range of actors nationally.

## LCR Netmapping Example



The result of these participatory network mapping exercises was twofold: (i) identification of actors that needed to be involved in the evaluation process directly; and (ii) greater understanding of the contextual dynamics at play that fed into the realist evaluation design.

To learn more about the evaluation, check out the [ECPAT website](#).

↪ *What level of participation of whom?*

Once all stakeholders have been mapped, the team then reflects on the purpose of including the most relevant in the evaluation process to define what form their participation should take. Here we provide a typology of four levels of participation and their responsibilities within the evaluation process for the most relevant stakeholders to be mapped on to.

Level of participation	Responsibilities
<b>High control over direction and resources</b>	<ul style="list-style-type: none"><li>• Define evaluation questions and scope</li><li>• Approve design and methodological choices</li><li>• Use findings for decision making</li><li>• Control resource allocation</li></ul>
<b>Direct involvement in evaluation implementation</b>	<ul style="list-style-type: none"><li>• Participate in specific methodological design</li><li>• Participate in data collection</li><li>• Participate in causal analysis</li><li>• Participate in development of recommendations</li></ul>
<b>Specialized input and guidance</b>	<ul style="list-style-type: none"><li>• Provide methodological expertise</li><li>• Provide contextual expertise</li><li>• Offer sector- specific insights</li></ul>
<b>Regular but limited participation</b>	<ul style="list-style-type: none"><li>• Share perspectives and experiences</li><li>• Offer feedback in key moments</li></ul>

*Authors own*

To ensure the value of **seeking greater equity** is driving decisions on who to include and in what form, the team needs to consider how power influences the ability of different stakeholders to engage meaningfully, striving to identify opportunities to create conditions for those normally excluded to have a seat at the table. This requires an examination of how different forms of power create barriers to participation. For more resources on how to engage with power see the [PowerCube](#) website. Here we provide one way in which we can understand how different forms of power can pose barriers to meaningful participation.

Forms of power	How power works	Barrier to participation
<b>Visible</b>	Power is the visible competition between interests in decision making	Some decide not to participate due to apathy, fear or lack of information and skills.
<b>Hidden</b>	Power is the ability to set the agenda of decision making	Some do not have the resources or skills to participate and cannot get their issues on the agenda.
<b>Invisible</b>	Power is the social norms and values that maintain the status quo	Social norms mean that certain people are not seen as legitimate participants or do not see themselves as having a right to participate. Or they accept the inevitability of an unjust and unequal system.

Adapted from [Participation in Economic Decision Making](#) (Oswald et al. 2018)

## Case Example

### From Narratives to Pathways: Participatory causal analysis in evaluating contributions to systems change in the context of Brazilian education

The case of a collaborative causal pathways evaluation with six grantees of Imaginable Futures in Brazil (published as a Causal Pathways case) illustrates three levels of engagement with different actors including the funder, grantees (partners) and participants in the interventions themselves. The context was a collective evaluation of how each partner contributes to shifting dynamics in the Brazilian education system to support greater equity.

#### Levels of evaluation and participating actors



In this [case](#), the evaluation team discussed with the funder how to navigate power across the funder-grantee interface which most commonly shuts down the space for open and honest reflection on what is and what is not working.

In this case recognition of power and careful facilitation by the evaluation team meant the implementing partners, represented by their leaders, could work alongside the funder to form a core team involved in all decision making. In this case they all had high control over the direction of the evaluation. A broader set of people from the partner organisations were involved only in the implementation itself, while project participants were involved to varying degrees - from respondents in some cases to more active agents in others (such as for example through collective sensemaking of data with participants).

### *Design governance and ensure quality in facilitation*

Based on the analysis of stakeholders and the level of participation desired in the evaluation, questions about how to govern and facilitate the causal pathways evaluation can be answered.

A governance body should be set up and include all actors who need to have high control over the direction and resources. Naming the governance actors, however, is not, on its own, sufficient to ensure **utilization** and **equity** - deciding who will facilitate the process in ways that support equity is also necessary. Here, quality of facilitation becomes the key condition for an equity focused deliberative process to govern the evaluation. Reflecting on core facilitator competencies is essential.

#### **Facilitator Competencies**

- **Inclusive Communications Skills:** the ability to create an environment where all actors feel welcome, respected, and empowered to share their perspectives. This requires active listening, using clear and accessible language, and ensuring that diverse voices are heard and valued.
- **Cultural Sensitivity and Contextual Understanding:** facilitator must demonstrate sensitivity to cultural nuances, power dynamics, and local contexts as relevant to the causal pathways being explored. This involves recognizing and mitigating potential biases, understanding the historical and social dynamics of the group, and adapting facilitation approaches to respect and honor local knowledge and traditions.
- **Reflexive and Adaptive Practice:** the ability to engage in continuous self-reflection and adaptability. This involves critically examining their own biases and assumptions; being transparent about their role and their limitations; an openness

to adjusting their methods based on participant feedback, and maintaining an attitude of humility and learning throughout.

- **Participatory Design Expertise:** have skills in designing participatory research and evaluation methods. This requires some knowledge of participatory tools and techniques and collaborative data collection methods. They must be able to support the creation of flexible, adaptive evaluation frameworks that center the experiences and insights of participants.
- **Systems Thinking and Complexity:** given the complexity of causal pathways and potentially different views about them within the group, the facilitator needs to understand interconnected systems, recognize and navigate power dynamics, help participants see broader contexts and relationships and synthesize diverse perspectives into coherent insights.
- **Ethical Facilitation:** maintaining ethical standards in research and evaluation is a critical skill. This requires the facilitator to have prior experience with research and evaluation ethics procedures such as confidentiality, informed consent but also ethics of care as is common in participatory evaluation.
- **Conflict Transformation and Mediation:** because a deliberate process including a diversity of actors can surface tensions it is helpful if the facilitator is able to create safe spaces for dialogue (not just name the need for them); see and use conflict as an opportunity for deeper understanding; be able to employ de-escalation techniques; focus on helping participants find common ground; manage disagreements constructively

*Authors own*

### *Agree appropriate causal evaluation questions*

Teams interested in undertaking a causal pathways evaluation should first establish that it is the right approach, by defining what causal questions are being asked. Given the value of embracing complexity that underpins a causal pathways approach to evaluation and learning, all questions that allow exploration of the following broad evaluation question could be considered.

## **Overarching Question for Causal Pathways Evaluations**

**How, why, for whom and under what conditions is a strategy/program contributing to intended or unintended (positive and negative) outcomes?**

Different types of causal questions could be asked within this overarching question, some of which are more directly about exploring causal pathways, others more



concerned with measuring impact. We list here all possible types of causal questions.

## Types of Causal Questions

### **Causal Process** Questions:

- What specific processes led to the observed outcomes?
- How did the intervention trigger change?
- What intermediate mechanisms connect the intervention to outcomes?

### **Causal Context** Questions:

- How do different contexts modify the causal mechanism?
- What contextual conditions enable or inhibit causal effects?
- How do local factors interact with intervention strategies?

### **Emergent Causality** Questions:

- How do unintended consequences emerge?
- What unexpected causal pathways developed?
- How do multiple factors interact to produce outcomes?

### **Comparative Causal** Questions:

- Which intervention strategy was most effective in creating change?
- How do different approaches compare in terms of causal impact?
- What are the differential effects of alternative interventions?

### **Equity Focused** Questions:

- How did causal effects differ across social groups?
- What variations in impact existed?
- Did the intervention affect different populations differently?

### **Systemic Causality** Questions:

- How did the intervention contribute to systemic changes?
- What broader transformations were triggered?
- How do multiple levels of the system interact causally?

### **Impact Contribution and Inquiry** Questions:

- To what extent did the intervention cause the observed changes?
- What difference did the program make compared to doing nothing?
- How much of the observed outcome can be directly attributed to the intervention?

*Authors own*

It is critical at this time to make clear that **a causal pathways approach is not ideal for answering impact contribution questions** that relate in particular to measuring the ‘net effect’ of an intervention against a counterfactual. If that is the intended goal then other designs and approaches may be more appropriate.

**Stage 2. Making evaluation design choices**

**The purpose** for this stage is to define the specific design, methods and tools that will be deployed to collect data and implement causal analysis as part of the evaluation. For causal pathways evaluations, the methodological choice stage is when we build a bricolage design that is **credible** and responds to the specific agreed causal questions.

At this stage, criteria are used to define specifically what rigor means to inform the choice of evaluation methods. Causal pathways evaluations employ methods that provide explanations of processes of change, and so tend to be [case-based](#), sometimes [theory based](#), and are underpinned by a variety of [causal inference](#) frameworks. Quality in design requires careful consideration of methods to respond to specific causal questions applying the following three criteria.

Criterion	How to apply to choice of methods
<b>Representation</b>	Consider how different methods and parts of methods enable engagement of priority groups such that their lived experiences help to explain causal dynamics
<b>Triangulation</b>	Consider how different methods and parts of methods can produce data from multiple sources and multiple lines of evidence
<b>Transferability</b>	Consider how different methods and parts of methods allow for contextual variation and exploration of the role of context as part of broader causal packages

*Authors own*

It is unlikely that one method alone will be sufficient to respond to the variety of causal questions being asked. Here, the practice of methodological [bricolage](#) is a way to design with intention, paying attention to our quality criteria.

## Case Example

### Evaluation of TeamUp

by Mieke Snijder & Tom Zwollo

TeamUp is a worldwide play and movement-based intervention developed by Save the Children, War Child, and the United Nations Children's Fund (UNICEF) to improve children's psychosocial wellbeing. TeamUp aims to be inclusive for all children aged six to eighteen years through its non-verbal modality - trained facilitators provide play, sports, and movement-based activities for children where each activity is connected to a psychosocial theme such as stress, friendship, or anger. Save the Children Netherlands commissioned the Institute of Development Studies (IDS) to undertake an impact evaluation of TeamUp in the Dutch asylum center context as part of an European Union-Asylum, Migration and Integration Fund (AMIF).

The evaluation called for a design that could work with the highly dynamic and largely unpredictable context of asylum reception centers. The evaluation employed contribution analysis, to respond to the main evaluation question: "How, why, and under which circumstances does TeamUp contribute to promoting the psychosocial wellbeing of children in Dutch asylum reception centers?" The commissioner's interest was to build a deep understanding of how TeamUp contributes to psychosocial wellbeing in different contexts, and when it does not. **During inception, the commissioner and evaluators agreed to use evidence quality rubrics to ensure a common approach to quality, appreciating that use of contribution analysis was a departure from previous experience.** Deliberations began with an explanation of seven criteria (independence, transparency, plausibility, uniqueness, triangulation, representativeness, transferability) chosen by the evaluation team as most relevant. After deliberation, the criteria of plausibility, uniqueness, triangulation, and representativeness were prioritized.

The evaluation team then drafted rubrics including descriptors for five performance levels for each criterion. The rubrics were applied during design to support methodological bricolage by making intentional decisions on how to combine methods to achieve quality. There was need to balance the value of participation and feasibility given the safeguarding concerns and challenging conditions for all participants in asylum centres. The table shows the resulting assessment of each method against the agreed quality criteria. In green we assessed that the method is likely to support the criterion, orange where it was less likely and red where it would not.

Method	Plausibility	Representativeness	Uniqueness
<b>Body mapping</b>	In order to understand how embodiment works and produce plausible narratives of change for children this method is critical.	The method supports deep exploration of children's experience, supporting this core group's representativeness in the evidence base.	This method is perhaps the most important for finding the 'magic spark' of TU given the embodiment is one of the features that makes it unique as an approach to psychosocial wellbeing and unlike other interventions in the AZCs.
<b>Children's data collection</b>	This method is not necessarily going to provide explanation of how change has happened in social-connectedness.	As a fun activity the children are likely to feel more involved in the evaluation supporting overall representativeness in the case study material.	This method is less likely to show unique TU contribution.
<b>Interviews with children</b>	The interviews are likely to help fill any gaps around plausibility by providing evaluator directed questioning.	This method will contribute the least to representativeness.	As a gap filling method this could support identifying uniqueness in an intentional way.
<b>Facilitator observations</b>	Complementing other methods the facilitator's observations could provide a richer picture for plausibility.	Representativeness of facilitator's interpretations of children's experience is not as important as children's views but will support use of findings.	Facilitators might have a sense of what the 'magic spark' is but will be from their own experience alone.
<b>Story collection and analysis</b>	The causal analysis will explicitly provide detailed information about the steps in the pathway to social connectedness. The high number of stories will add to the plausibility of these claims.	Involving children in the analysis will provide strong representativeness to the evidence generated.	Unlikely to reveal uniqueness in a specific way because the question it responds to is broader than TU by definition.
<b>Interviews with facilitators, COA and RCs</b>	Realist interviews will support plausibility by explicitly seeking to fill gaps in the causal pathways uncovered through the other methods.	This is the least representative method of all as it <u>an</u> evaluator driven exercise.	Realist interviews can dig deep to identify uniqueness if it remains opaque through other methods.
<b>Most Significant Change with parents</b>	Including a causal analysis within the MSC process will provide useful details for plausibility.	As the only method that will work with parents explicitly it will provide a particular aspect of representativeness that no other method can.	If uniqueness is informing the analysis process - so seeking to understand contribution and probing around TU, then the method could support this criterion (this could drive the analysis away from the participant driven definition of significance).

As a result of this assessment, the evaluation team adapted the design in several ways to support greater quality. For example, the Most Significant Change (MSC) with parents was adapted to bolster plausibility and uniqueness by taking elements from Outcome Harvesting which ask participants to detail how the intervention is contributing to the change identified as most significant. To support triangulation, the team decided to prioritize collecting data from the children of the parents who were involved in the MSC. The team included a new method of children's research activities as a way to deepen their engagement in exploring the immediate effects of TeamUp on processes. Using research booklets, children explored their own questions as well as reflected on TeamUp and how they build friendships. Further, body mapping before and after the TeamUp sessions, allowed children to share what they experienced in their bodies and reflect if and why these changed.

The full report which includes final assessment using the evidence rubrics can be [accessed via](#) the Save the Children repository.

In Stage 4 we then describe how these same criteria can be used to develop rubrics to assess the strength of the causal claims that result from the evaluation.

### Stage 3. Implementing causal analysis

**The purpose** of this stage is to apply the chosen evaluation methods in ways that support quality causal analysis.

Some of the methods that are commonly used in causal pathways evaluations include within them explicit steps or processes for quality assessment. For example, [Process Tracing](#) includes a series of ‘tests’ that allow the evaluator to assess the strength and relevance of data and evidence to test specific causal hypotheses (see case example box).

#### *Case Example*

#### **CocoaLife Evaluation in Côte d’Ivoire**

An independent evaluator, CARE International UK’s MEL team, CARE Côte d’Ivoire and their partners 2A carried out a participatory evaluation of the CocoaLife project which aimed to improve the livelihoods of 424 communities working in coca value chains. The evaluation team employed an adaptation of process tracing, called contribution rubrics, which combined theories of change, outcome statements, evidence tests, and rubrics.

As the project did not have a detailed theory of change, the evaluation team developed this in a workshop and identified several contribution claims. The workshop also provided a training on process tracing for CARE UK’s MEL team, the project team and their partners. One key area of work was supporting communities to establish Community Development Committees (CDCOMs) and enabling them to develop Community Action Plans (CAPs). The outcome identified to which the project claimed to have contributed was ‘CDCOMs influence the provision of selected essential infrastructure, enabling co-financing from cooperatives, communities, and other actors.’ The team elaborated three causal chains for the community-based work:

1. Community advocacy;
2. Community resource mobilization, and;
3. Cocoa Life convening and brokering.

The chains were broken down into 25 components (or steps), with two complementary components from the donor, Mondelēz International, through the creation of the “Opportunity Fund” which co-financed infrastructure. Specific cases were identified to substantiate contributions to outcomes. These aligned with the most common priorities identified by communities in action plans.

These aligned with the most common priorities identified by communities in action plans. These included the construction of a health centre in the terroir of Sikaboutou and a water pump in the terroir of Gozon (there are usually 8-9 villages which form a terroir cluster).

Using process tracing evidence tests, the project team and their partners then identified necessary evidence they would expect to find (hoop test) and hope to find (smoking gun) if their explanation were accurate and raise confidence in the validity of contribution claims. This was essentially a reasoning exercise regarding the “probative value” of evidence and what finding it (or not) might mean for the team’s contribution claims. In order to trace the whole process, the evaluation team reviewed available secondary evidence for every step and the project team gathered additional primary data on most of these steps.

In Sikaboutou, for example, action plan records revealed that 21 problems were identified, including the construction of a health centre. Interviews and account data revealed that communities mobilised resources and the Coffee and Cacao Council (CCC) also provided funds for the construction of the health centre, and photos demonstrated that the new health centre had been constructed. In addition, unprompted by the project team, the evaluator found a video on Extrait RTI television’s YouTube channel on the 10 May 2019 which showed the inauguration of the health centre. The video demonstrated that the construction was due to funds from Mondelēz, the Coffee and Cocoa board and community contribution. The project manager was also clearly visible in the video and participating actively in the event. This evidence gave credibility to the contribution claims which were appraised using rubrics.

Altogether, the evaluation demonstrated how project teams and their partners can be actively involved in theorising, assessing evidence, and collecting evidence. Independent evaluators can also play a helpful role to guide the process and offer methodological quality assurance.

Further information about the case [here](#).

Other methods have other ways of defining what rigor means. All methods should be applied in appropriate ways. Here, we encourage evaluation teams to consider all approaches and methods included in the list on the [Causal Pathways Resource Hub](#). Other useful guides that provide overviews and details of different methods and approaches that are also useful include:

- The UK Government [Magenta Book](#)

- The World Bank's [IEG Method Paper Series](#)
- The LIEPP [Methods Briefs](#) series
- The [EvalParticipativa](#) site
- The [Participatory Methods](#) site

Central to using these methods **within a causal pathways evaluation** is how 'causal analysis' is implemented. Quality here rests on the ability of evaluation teams and those directly involved in the causal analysis of data to employ **reflexivity** and **critical reasoning** as the causal pathway is interrogated. As [Lynn and Apgar](#) discuss, it is important to safeguard against constructing scenarios to make the data 'fit' rather than critically examining what the data is telling us.

**The practice** of critical reasoning in causal analysis involves:

- ❖ **Identifying causal hypotheses and interrogating each step within a causal chain** - many of the theory based evaluation methods that are useful in a causal pathways evaluation, begin by first hypothesizing how a pathway might work. In some methods, it is necessary to then zoom in to unpack specific causal links, for example the practice of defining [causal hotspots](#) in contribution analysis helps to interrogate specific steps in the causal chain that are critical for explaining how the causal pathway unfolds.
- ❖ **Coding for causality** - moving beyond hypothesizing how change might happen or has happened to examining how it has happened, requires that we analyse data specifically looking for cause and effect relationships. The causal map software is one approach to this - there are several examples of how this causal coding was done within QuIP evaluations developed as [case studies](#).
- ❖ **Testing causal relationships and exploring rival explanations** - testing causal relationships goes beyond identifying or describing them (coding for causality) to inquire into them through seeking both confirming and disconfirming evidence. In particular, seeking [disconfirming evidence](#) that can challenge causal hypotheses makes for more robust explanations. Exploring rival explanations is another way to strengthen the extent to which causal links are being tested rather than simply evidenced.
- ❖ **Practicing triangulation** - triangulation tests the consistency of findings obtained through different means and increases the chance to assess how multiple causes influence results. There are at least six different ways which triangulation can be implemented within causal analysis: (1) seek to



triangulate data from different time periods, locations, sources (eg. multiple interviewees, multiple questionnaires) and lines of evidence (i.e., different source types interviews and questionnaires); (2) include more than one evaluator or analyst, helping to provide multiple interpretations; (3) use multiple theories to help frame and explore specific causal links; (4) using multiple methods can build confidence in results. The extent to which all forms of triangulation can be practiced will depend on the shape and size of the causal pathways evaluation.

### *Case Example*

#### **From Narratives to Pathways: Participatory causal analysis in evaluating contributions to systems change in the context of Brazilian education**

The case of a collaborative evaluation with six grantees of Imaginable Futures in Brazil (published as a Causal Pathways [case](#)) illustrates how critical reasoning was used when analysing narratives of change collected within an adaptation of Outcome Harvesting.

“Mapping the causal pathways related to outcomes of each collected story was a way to make visible aspects that needed further substantiation, which entailed a new data collection moment, whether going back to the same actors or speaking to new actors. The substantiation needs varied among confirming contribution claims (“is it possible to conclude that this action of our program contributed to that aspect of the narrated change?”); better understanding the previous context in the participant's life before they joined the program (“do we really understand intrinsic motivations of the participant to situate this story in a richer context?”); and further inquiring into the impacts of the individual changes that had been narrated by participants (“what changes had the individual changes actors told stories about enabled, and did those have further ripples out into the system?”). This also led to the teams identifying aspects that could be strengthened by triangulation of data, in this case listening to the points of view of other actors in the system about that collected story to confirm, build confidence in or provide a distinct position on what was being claimed.” ([CP case study](#))



## Stage 4. Assessing the strength of evidence of causal claims

The result of applying fit-for-purpose methods and causal analysis of data is **a set of claims about how the causal pathway has evolved**, and where possible, inference is made on what contributed to outcomes in the pathway. **The purpose** of this stage is to assess how strong the “probative value” of the evidence is - to ask what makes a particular explanation better or worse ([Ribeiro, 2019](#)). In the case of causal pathways evaluation we are most concerned with the **strength of causal explanations**.

In practice this might already be embedded in the method of use (e.g. as explained in Stage 3 guidance with the example of Process Tracing), but often within a causal pathways evaluation, evaluation teams are building causal claims across methods used in a bricolage design. In these cases, many evaluators are now developing and applying strength of evidence rubrics to help make judgements collectively. For more on the many different kinds of [rubrics](#) in use follow [Julian King](#)'s work.

### *Case Example*

#### **Use of Evidence Rubrics in the CLARISSA program**

In the CLARISSA program (see full case study write up [here](#)) causal analysis was embedded within an adaptation of the Outcome Harvesting methodology as one method used within the contribution analysis design. Details of how evidence rubrics were developed and deployed to support this causal analysis is described in [this paper](#). In summary, the CLARISSA [quality of evidence rubrics](#) were applied by the evaluation and implementation team to intentionally reflect on how strong the existing evidence was in explaining the causal pathway and to identify where gaps existed. By layering critical reflection on plausibility, triangulation, representativeness and uniqueness, the team was able to design the substantiation step for an external evaluator to critically examine the causal claims the program was making.

Here are some tips for developing and working with strength of evidence rubrics:

- ❖ **You don't have to start from scratch** - see [here](#) for a set developed for case based methods which can be a useful starting point.

- ❖ **You do have to contextualise them** - the power of rubrics is that they are a flexible tool that can be adapted to suit specific contextual needs. Take time to contextualise them with all evaluation stakeholders who can provide insights on the contextual conditions that might influence how a causal claim is made.
- ❖ **Decide on how many levels of performance are useful** - as a rule of thumb, we recommend no less than three and no more than five levels.
- ❖ **Apply them through a facilitated process** - rubrics are particularly good at creating the space for debate and deliberation when boundaries are fuzzy, there will be different interpretations of what 'good enough' triangulation means or, how many rival explanations needed to be discarded. This debate will be most fruitful if there is strong facilitation.
- ❖ **Make use of rubrics transparent** - we recommend that the best way to show the quality of the causal pathways work is to be transparent about how rubrics were developed and used. Including final rubric assessment with the reasoning behind the assessment in evaluation reports can build this transparency.

## References

Apgar, M., Aston, T., Snijder, M., & Zwollo, T. (2024). Raising the Bar: Improving How to Assess Evidence Quality in Evaluating Systems-Change Efforts. *The Foundation Review*, 16(2).

<https://doi.org/10.9707/1944-5660.1712>

Apgar, M., Snijder, M., Paul, S., Ton, G., Prieto Martín, P., Veitch, H., et al. (2024). Evaluating CLARISSA: Evidence, Learning, and Practice. The Institute of Development Studies and Partner Organisations. Report. <https://doi.org/10.19088/CLARISSA.2024.050>

Bertermann, K., & Martin, A. (2024). Philanthropy Needs More Disconfirmation Bias. *Stanford Social Innovation Review*. <https://doi.org/10.48558/2GNY-1171>

CLARISSA (2023) 'CLARISSA's Quality Of Evidence Rubrics', Design Note 2, Brighton: Institute of Development Studies, DOI: 10.19088/CLARISSA.2023.003

Clark, L. & Apgar, M. (2019). Unpacking the Impact of International Development: Resource Guide 2. Seven Steps to a Theory of Change. The Institute of Development Studies and Partner Organisations. Online resource. <https://hdl.handle.net/20.500.12413/14793>

H M Treasury. (2020). The Magenta Book: Guidance for evaluation. UK Government. Retrieved from: <https://www.gov.uk/government/publications/the-magenta-book>

Lynn, J. & Apgar, M. (2024) Exploring causal pathways amid complexity: Understanding when and how causality can be made visible in Newcomer, K., & Mumford, S., Research Handbook on Programme Evaluation. Edward Elgar Publishing Ltd.: <https://www.elgaronline.com/edcollchap-oa/book/9781803928289/book-part-9781803928289-25.xml>

Olmos-Vega, F. M., Stalmeijer, R. E., Varpio, L., & Kahlke, R. (2022). A practical guide to reflexivity in qualitative research: AMEE Guide No. 149. Medical Teacher, 45(3), 241–251. <https://doi.org/10.1080/0142159X.2022.2057287>

Oswald, K., Apgar, M., Thorpe, J., & Gaventa, J., (2018). Participation in Economic Decision-Making: A Primer. The Institute of Development Studies and Partner Organisations. Online resource. <https://hdl.handle.net/20.500.12413/13815>

Ribeiro, G. (2019). Relevance, probative value, and explanatory considerations. The International Journal of Evidence & Proof, 23(1-2), 107-113. <https://doi.org/10.1177/1365712718816740>

Schiffer, E., & Hauck, J. (2010). Net-map: Collecting social network data and facilitating network learning through participatory influence network mapping. Field Methods, 22(3), 231–249. <https://doi.org/10.1177/1525822X10374798>

Snijder, M; Apgar, M. (2024) Psychosocial wellbeing for children in Dutch asylum reception centres. TeamUp Impact evaluation. Final Evaluation Report. April 2024. Institute of Development Studies. Accessed: <https://resourcecentre.savethechildren.net/pdf/Final-impact-evaluation-report-with-recommendations--April-2024.pdf/>

van Draanen, J. (2017). Introducing Reflexivity to Evaluation Practice: An In-Depth Case Study. American Journal of Evaluation, 38(3), 360-375. <https://doi.org/10.1177/1098214016668401>