CDI Centre for Development Impact
PRACTICE PAPER

Innovation and learning in impact evaluation

The Art and Craft of Bricolage in Evaluation

Abstract This CDI Practice Paper by Tom Aston and Marina Apgar makes the case for 'bricolage' in complexity-aware and qualitative evaluation methods. It provides a framework based on a review of 33 methods to support evaluators to be more intentional about bricolage and to **combine** the component parts of relevant methods more effectively. It discusses two cases from practice to illustrate the value added of taking a more intentional approach. It further argues that navigating different forms of power is a critical skill for bricolage, and that doing so can help to ensure rigour.

1 Introduction

Over the past decade, various efforts have been made to determine how to choose fit-for-purpose methods and tools for evaluation (Stern *et al.* 2012; Befani 2016). Befani (2016: 3) notes that 'it is now largely accepted that a wide range of methodological options are appropriate, under different circumstances, to evaluate the impact of development programmes'. And yet, the range of potentially appropriate methods has created a selection problem for evaluators, programme implementers, and evaluation commissioners.

Rather than any single 'gold standard' method, it is argued that the 'best' method is dependent on what questions an evaluation asks, the attributes of the intervention being evaluated, and available designs linked to the intended uses of the evaluation (Stern *et al.* 2012; HM Treasury 2020). This so-called 'design' triangle and subsequent revisions such as the 'choice' triangle, which calls attention to additional evaluation goals and methodological requirements, are all helpful in making appropriate choices (Befani 2020). However, guidance still tends to assume that the evaluator must make a choice between methods and remains silent on how to combine them.

Rather than supporting evaluators to choose **between** methods, this practice paper aims to support evaluators to effectively **combine** the component parts of methods

through 'bricolage', focusing on complexity-aware and qualitative methods. Evaluation methods build on the foundations of others. Methodological recombination has often strengthened the quality of evaluations, enabling evaluators to take the best bits of some methods to bolster the weaknesses of others. We offer criteria for rigorous qualitative methodological bricolage; we then define core functions from appropriate methods and provide examples of bricolage from practice. Finally, we provide some thoughts on how bricolage can be negotiated as part of evaluation when seen as a participatory multi-stakeholder process.

2 What is bricolage?

Bricolage was first coined by French anthropologist Claude Lévi-Strauss, derived from the French verb *bricoler* ('to tinker') and linked to the term *bricoleur* which means 'a jack-of-all-trades'. According to Lévi-Strauss (1968), bricolage is an attempt to reuse a heterogeneous repertoire of available materials to solve new problems.

In qualitative research, the notion of bricolage has long been understood as a strategy that adds rigour (Denzin and Lincoln 1999; Chambers 2015). In practice, although many evaluators are *bricoleurs*, the dogma of evaluation 'brands' and notions of a supposed 'gold standard' have until recently hindered bricolage from gaining attention in the evaluation field (Patton 2019; Hargreaves 2021; Aston *et al.* 2021).

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. This is what we refer to as bricolage.

3 The case for qualitative bricolage

Evaluation methods have different strengths and weaknesses, so it is common to combine them to help answer different evaluation questions or respond to specific links in a theory of change. Yet, the emphasis has typically been on qualitative methods filling the gaps of quantitative methods (Jimenez *et al.* 2018). The discussion has tended to focus on the distinction between quantitative methods employing a statistical or counterfactual logic; and qualitative (or case-based) methods employing configurational or generative (i.e. theory-based) logics (Stern *et al.* 2012). The consequence has been to reinforce an unhelpful dichotomy that fails to appreciate the more nuanced ways in which combining various qualitative methods can respond to conditions of complexity in evaluation. Here we focus on how bricolage of these methods can add to the momentum of broadening appropriate evaluation designs.

Programmes are increasingly operating under conditions of complexity, addressing 'problems' that are the result of multiple interacting and often hidden dynamics in particular contexts. Such interventions require multiple strategies and engagement with diverse actors in systems, leading to unpredictability in outcome pathways. Both intended and emergent outcomes need to be captured, calling for evaluation approaches that support iterative learning as change happens (Apgar, Hernandez and Ton 2020). This trend has matured in recent years, from conceptual and metaphorical use of the term 'complexity' to empirical discussions about methodological implications in more nuanced and specific ways. The result is an appreciation of the need to combine and mix methods and approaches rather than choosing the 'best' method *a priori* (Hargreaves 2021).

Evaluations engaging with complexity and supporting learning are often theory based. They aim to explain how, why, and for whom interventions work (Rogers and Weiss 2007). Theory-based approaches tend to be more concerned with contribution and levels of confidence rather than sole attribution. Such approaches appreciate that there is commonly a conjunction of contextual and intervention factors that together produce an outcome. As well as being theory based, they tend to be case centred, focused firstly on within-case explanations – what is known as 'causes of effects'. They 'often start

with events that occurred in the real world – observable outcomes – and work backwards to ask about causes and search for combinations of conditions which lead to the identified outcome' (Goertz and Mahoney 2012: 42). They can support explanations that embrace social, political, and environmental contexts as vital to understanding complex processes of change.

It is possible to adopt certain parts of methods and combine them to better appreciate context as an intrinsic part of the causal process. Maxwell (2012) argues that not only do causal relationships potentially differ across contexts, but context is also an intrinsic part of the causal process (Maxwell 2004: 6). This interaction is key to explaining how and why an intervention works. Methods such as Realist Evaluation (RE) and Process Tracing (PT), for example, consider the interaction between contexts and mechanisms at the heart of our inquiry. It is important to specify the contextual conditions that must be present for a mechanism to be triggered and for an outcome to happen. As Greenhalgh and Manzano (2021: 2) argue, 'context [should be] conceptualised as relational and dynamic features that shape the mechanisms through which the intervention works'.

A commonly overlooked aspect of context in evaluation is the experience of the central actors of the processes of change – the 'stakeholders' themselves who are the targets of social change efforts. Participatory methods centre people's experiential and embodied knowledge and can be used to explore the meaning of social reality of those enmeshed in context (Heron and Reason 1997; Burns, Howard and Ospina 2021). A recent exploration of well-known participatory evaluation methods – including Most Significant Change (MSC), Outcome Mapping (OM), Ripple Effects Mapping (REM), SenseMaker, and the Reality Check Approach (RCA) – suggests that these are particularly valuable for bringing lived experience into our understanding of how outcomes are achieved in context (Apgar and Allen 2021).

Most use narratives of change; MSC and REM approaches may look at collecting and working with up to a few hundred stories, while SenseMaker is designed to work with hundreds and even thousands of stories. They allow us to go deeper and understand rich experience and causal pathways in context. The co-construction of knowledge that is assumed within participatory methods, however, does not fit neatly into a view of evaluation as supporting external and unbiased views that should ultimately judge performance. These methods share the criticism of small-N approaches using qualitative, case-based methods, such as confirmation bias and courtesy bias. These risks notwithstanding, there is room for greater appreciation of how they might enhance rigour by increasing the level of triangulation of sources and quality of critical thinking, as well as design relevance (Preskill and Lynn 2016).

4 Maximising strengths and bolstering weaknesses

How we choose which combinations of methods are appropriate should rely on what our evaluation questions are, and ultimately what evaluation commissioners, evaluators, and programme managers value. Critical to combining is a need to look across methods (or parts of them) to ascertain how robust conclusions are likely to be from any particular mixture. Aston *et al.* (2021) argue that for complexity-aware and learning-oriented programming, rigour can best be understood through the following criteria:

- **Reasoning:** critical thinking is fundamental to evaluative reasoning. This may include consideration of alternative explanations and interpretations, and a search for outliers.

- **Credibility:** credibility and the degree of confidence in findings speak to concerns of internal validity and the distinctiveness of effect patterns related to evidence to support contribution claims in context.

- **Responsiveness:** questions, methods and analyses should reflect local stakeholders' values and cultural context, and be sensitive to their experiences and definitions of success.

- **Utilisation:** the quality of the learning process, actionable evidence, and related use of evaluation findings are fundamental.

- **Transferability:** as context matters, transferability and reflection on other moderating factors is needed. This relates to external validity, with an emphasis on how the context enables intervention outcomes.

Redefining rigour in this way better aligns with more complexity-aware, participatory and context-sensitive logics. Yet, no single method covers all these criteria adequately, providing a strong justification for bricolage. Methods such as PT, for example, may be strong on reasoning and credibility because they have specific tools for this (e.g. evidence tests and explicitly assessing rival hypotheses). But they may be weaker for other criteria such as responsiveness and utilisation where alternative methods such as MSC or Outcome Harvesting (OH) are stronger. Rather than focusing on methodological weaknesses (Ton 2012), we emphasise the strengths of particular steps, activities, or tools in methods that could be used to bolster the weaknesses of others.

5 Recombination through bricolage

We assessed 33 relevant approaches and methods, and identified common trends and the value added of particular activities and tools offered (see Annexe 1) – what we call **functions**. In Table 1, we present ten core functions and

illustrate how particular steps or tools in relevant methods fit with the above rigour criteria.

In practice, evaluators often skip steps within methods for pragmatic reasons such as time and resource limitations. Further, certain steps, activities, or tools within methodologies may also be implicit, poorly conceived, or deemed superfluous. Evaluators rarely employ all formal steps in MSC or OM, for example. In MSC, there are arguably four key steps: (1) deciding the types of stories to be collected; (2) collecting stories; (3) determining which are the most significant; and (4) sharing the stories and discussion of values with stakeholders. Other steps are often considered optional. Similarly, it is relatively rare in OM to articulate organisational practices or fill out strategy or performance journals, as these are not seen to be necessary in all circumstances. So, skipping activities or tools may not be a weakness, and substituting better developed steps from other methods can strengthen them.

Evaluators tend to use favoured activities and tools from their preferred methods alongside the primary method chosen for a particular evaluation. For instance, it is relatively common to skip the substantiation step in OH, so evaluators may rely on steps in other methods to triangulate or validate to increase their credibility (e.g. PT evidence tests). Taken together, the functions and rigour criteria shown in Table 1 below can support more intentional combining, not solely of whole methods but of parts of methods.

Relatively few evaluation methods take account of context analysis **explicitly** as part of explaining change processes. Given the weakness of many methods in making context analysis explicit, evaluators fill this gap through political economy analysis or other forms of power and problem analysis. Alternatively, RE makes context a central concern in the development of context, mechanism, outcome statements.

Developing some form of outcome pathways is central to most theory-based methods. Theories of change are a common foundation for many evaluations that subscribe to a qualitative-causal culture. Contribution Analysis (CA), for example, uses the development of a theory of change as an anchor and employs aspects of other methods (e.g. PT evidence tests) within this general framework.

Appreciative inquiry is a common approach within participatory methods such as REM and MSC, or adaptations of MSC such as Collaborative Outcomes Reporting (COR). Articulating and appraising outcomes is central to methods such as OH, which typically supports project teams (and other evaluation stakeholders) to narrate outcomes and assess the significance of these outcomes as part of sampling as well as analysis.

Table 1 How functions support rigour criteria through methods

Function	Connection to rigour criteria
Context analysis of potentially contributing factors and conditions	Provides boundaries for causal mechanisms, enhances external validity of evaluative judgements, and supports transferability (e.g. context, mechanism, outcome configurations in Realist Evaluation).
Developing outcome pathways for causal processes	Helps to structure evaluative reasoning, enhances evaluability, and can bolster credibility of findings. If developed through a participatory process, this can contribute to responsiveness and utilisation (e.g. theory of change in Contribution Analysis, or developing causal chains in Process Tracing).
Appreciative inquiry of stakeholders to begin with an understanding of what matters to them, how they envision change, and what they want to learn about	Helps to ensure that evaluation is relevant and representative of stakeholders' experiences, and supports utilisation when orienting evaluation to learn from and about these experiences (e.g. collect significant change stories in Most Significant Change, or social inquiry in Collaborative Outcomes Reporting).
Articulating outcome narratives to explain outcomes and contributions	Enhances evaluative reasoning and the testability of effect patterns, which raises credibility, and can increase responsiveness if these are developed in a participatory way (e.g. draft outcome statements in Outcome Harvesting or progress markers in Outcome Mapping).
Appraising significance of outcomes to explain why a change is important	Enhances evaluative reasoning and improves responsiveness if appraisal is a participatory process (e.g. collecting significant change stories in Most Significant Change or collecting significance data for outcome statements in Outcome Harvesting).
Iterative sensemaking of hypotheses and contribution claims	Helps to strengthen reasoning and enhance credibility and can also be conducted in a participatory way to enhance responsiveness (e.g. assessing a contribution story and seeking additional evidence in Contribution Analysis or revising the conceptual model in Multiple Lines and Levels of Evidence).
Testing strength of evidence underpinning contribution claims	Appraising evidence strength stress-tests evaluative reasoning. It strengthens internal validity and thus increases the credibility of findings and evaluative judgements (e.g. evidence tests and rival hypothesis assessment in Process Tracing; ruling out possible alternative explanations in General Elimination Methodology).
Validation of outcomes with communities, peers and/or experts	Strengthens the credibility of evaluative judgements (internal validity). If conducted in a participatory way, it can contribute to responsiveness (e.g. outcome panel in Collaborative Outcomes Reporting; contribution trial in Process Tracing with Bayesian Updating).
Causal pattern comparison between outcomes	Can help identify trends and outliers, thereby contributing to evaluative reasoning and transferability of findings (e.g. data matrix and truth tables in Qualitative Comparative Analysis; SenseMaker's visualisation of patterns).
Supporting utilisation of evaluation findings to inform future programming	Enables use of evaluation findings and supports transferability through adapting programming and informing new programming (e.g. supports use in Outcome Harvesting).

Several methods are designed to include iterative sensemaking, such as the assessment of a contribution story in CA, RE's refinement of programme theories to demonstrate common mechanisms of change, or SenseMaker's efforts to visualise and make sense of patterns. We also find several methods that are concerned with testing the strength of evidence and credibility of explanations such as PT via evidence tests, outcome panels in COR, and member checks in REM.

Methods such as Qualitative Comparative Analysis (QCA) are explicitly geared towards making comparisons between cases via truth tables. But other methods such as MSC and SenseMaker also provide an opportunity to compare patterns against key domains of change. Finally, while all evaluation methods should take evaluation use seriously, only a handful of the methods reviewed (e.g. OH) make this function explicit.

6 Examples of bricolage in practice

Evaluation design is a craft that requires evaluators to master multiple techniques that may stem from various methods. Evaluators also substitute the steps of some methods for others because they are deemed a better fit or have advantages for a particular evaluative exercise. Two examples from our practice serve to illustrate 'how to bricolage' as a **process** over time. We present here our retrospective interpretation of how bricolage happened. The examples were not implemented in a formal and structured way, as expressed above. The reality of most evaluations is not as neat linear processes, but rather as messy emergent and iterative processes. We share these examples to make real what bricolage looked like in practice, and then compare these choices against the framework to motivate more intentional practice.

Bricolage in PERL

The Partnership to Engage, Reform and Learn (PERL) programme in Nigeria shows how bricolage can be layered over time. PERL's monitoring and evaluation (M&E) team built on the experience of the State Accountability and Voice Initiative (SAVI) programme. SAVI evidenced the contribution to outcomes of government responsiveness to citizens through a blended MSC and OH template, which was combined with rubrics to assess levels of performance. Outcomes were validated via a 'peer review' function by local independent impact monitoring groups ('critical friends' of SAVI) – and by neighbouring state teams and partners (Derbyshire *et al.* 2016).

PERL's M&E team built on these foundations and added an explicit selection step to shortlist outcomes to validate. It developed a higher-level theory of change with ten general hypotheses around which teams would harvest evidence annually and write up case studies. PERL's learning partner, the Overseas Development Institute (ODI), periodically carried out deep-dive case studies, some of which would employ CA. Focused on testing hypotheses in the theory of change (e.g. value added of regional learning hubs) and drawing on MSCs as evidence sources, ODI researchers incorporated Contribution Rubrics' significance, contribution and strength of evidence criteria in a CA evaluation to structure evaluative judgements.

After several years gathering case study evidence, the Foreign, Commonwealth & Development Office urged PERL to re-evaluate the significance of previous years' changes based on criteria such as equity, sustainability and scale, as well as to reassess the rigour of PERL's MSC-OH hybrid. This led to a call to introduce PT as part of case-study development, to adapt scales and rubrics for significance scoring, and to set up an MSC task team, with internal peer reviewers from different organisations in the programme validating the credibility of outcome narratives.

PT was primarily designed to better explain narratives of change over time since the beginning of the programme and to strengthen contribution claims through the logic of evidence tests. PERL also included a checklist for MSC authors to strengthen contribution claims, including strength of evidence criteria such as specificity, triangulation and transparency, and explicitly comparing claims with alternative explanations and interpretations. The Foreign, Commonwealth & Development Office commissioned an independent 20-year study to help put changes over time into perspective. ODI's study built on the MSCs as key data sources. As a key focus was on identifying contextual enablers and mechanisms of change in the education and health sectors, it combined realist synthesis and PT in particular states (Piron and Waldock 2021).

Bricolage in WorldFish

WorldFish's aquatic agricultural systems programme in Bangladesh, Cambodia, the Solomon Islands, the Philippines, and Zambia provides another example of bricolage through evolving programme M&E design for both accountability and learning (Douthwaite, Apgar and Crissman 2014). This was a large agricultural systems programme that aimed to put research at the service of addressing development challenges through a research-in-development approach (Douthwaite *et al.* 2017). The central evaluation question was related to the research approach, and focused the M&E system on understanding if and how Participatory Action Research (PAR) and multi-stakeholder engagement contributes to improved livelihoods for poor and marginalised people.

In the five aquatic agricultural systems, a participatory design and diagnosis team that included key stakeholders of the fisheries and livelihoods systems (ministries of agriculture and/or fisheries, national agricultural research centres, community leaders, etc.) undertook context analysis. The team's remit was to define the focus of interventions that respond to locally defined visions of success (articulated through the Community Life Competency Process). The participatory design and project set-up culminated in a stakeholder workshop modelled on the Participatory Impact Pathway Analysis (PIPA) methodology (Apgar and Douthwaite 2021) to agree on the broad outcome pathways for research and development interventions to pursue. Consequently, the programme's monitoring system focused on tracking progress along these outcome pathways. OH was the overarching approach used to build real-time understanding of change and develop the contribution claims to be the focus of the impact evaluation phase. However, in the M&E team's experience, the substantiation step of OH was often poorly implemented and lacked the required focus on developing contribution claims.

Responding to this concern about OH, the team decided to bricolage MSC and use CA. Outcomes were identified

based on existing documentation of change through: documented significant stories of change (using an MSC process in some sites where the implementation team periodically collected stories, with stakeholders involved in defining significance) and ongoing documentation of work on the ground (an embedded PAR documentation system). The M&E team then described outcomes using an agreed template, which included sections on significance and contribution. Next, in a participatory workshop, the implementation team and – in some cases – other stakeholders (those involved in the earlier PIPA process) clustered the outcome descriptions into broad areas of change. For each cluster, multiple-cause diagrams were used to develop a theory of change working backwards from the outcome. This resulted in identifiable outcome trajectories of patterns of change in key actors – both partners and community – with clear contribution claims from the programmatic approach. The evidencing step focused on the contribution claims related to the programme's research-in-development approach. This first required a process to decide on which outcome stories within each of the emerging trajectories was 'most significant' for partners involved in the programme.

Gathering the evidence of the contribution of the programmatic approach across all five sites led to the development of an overarching theory of change of how research in development works. Building on this evidence, a design for a realist explanation was developed for the impact evaluation phase (Douthwaite *et al.* 2017). This form of bricolage is particularly important in the context of participatory programming when it is not possible to have *a priori* a causal theory of change. In these cases, methods such as MSC and OH used as part of programme monitoring can help to build causal understanding of outcome pathways.

The two cases demonstrate how formative and summative evaluations recognised the benefits of several key functions of another evaluation method, but not the method in full. These were considered to have specific advantages for a particular task and bolstered the weaknesses of the other methods chosen. The PERL example shows the benefits of several functions of bricolage: context analysis, particularly in the realist synthesis, underpinned by political economy analysis; articulating outcomes via PT narratives of up to 20 years; appraising significance of outcomes via MSC long- and shortlists; testing strength of evidence via the logic of PT evidence tests combined with rubrics to assess levels of performance; validating outcomes through different layers of MSC peer review; and comparing causal patterns via meso-level realist mechanisms and testing the overarching theory of change propositions. The WorldFish example shows the benefit of including stakeholders in context analysis and developing outcome pathways at the outset using PIPA; articulating outcomes via adaptations of OH and combining with MSC to appraise their significance;

iterative sensemaking through participatory workshops as part of OH; and outcome validation through adding contribution evidencing. And it shows how all these steps together can inform CA of a higher-level theory of change.

7 Negotiation: navigating power

Making choices about evaluation questions, design, and methodological recombination is not solely the realm of the evaluator, nor is it solely a technical challenge. In practice, bricolage has to respond to the goals of different evaluation stakeholders with diverse perspectives, positions, and power over decision-making. Bricolage, therefore, is not simply a technical challenge (the **craft**), but a relational and political one (Roche 2015). The **art** of bricolage is to recognise that different stakeholders, evaluators included, have different degrees and forms of power, and to navigate methodological decisions and their implications with this in mind.

Commissioners have the power to set the parameters for evaluations – they can open or shut down space for particular methods. We see an increasing opening for theory-based, complexity-aware and context-sensitive inquiry by enlightened donors, as the PERL case demonstrates (Piron and Waldock 2021), despite a wider preference in the international development sector for context- and complexity-insensitive quantitative evaluation designs. Programme implementers, however, have the power to feed evidence back into implementation, yet might prefer to have simple dashboards based on predefined indicators and results frameworks that hide the nuance and complexity of outcome pathways to simplify decision-making. The current move towards localising and decolonising development is creating new momentum for ensuring the people, groups, and communities that programmes being evaluated aim to serve are included throughout the programme cycle, from planning and design all the way through to evaluation. The needs and values of these groups are unlikely to align neatly with those of other stakeholders.

Navigating these different forms of power is a critical skill for bricolage and fundamental to build credibility, responsiveness and utilisation, in particular. Taking a participatory approach from the outset can be helpful because it centres questions on whose needs, experiences, and knowledge should be included and how. In our discussion of the ten functions and ways in which different methods are combined, we can see how participation within functions – such as using participatory methods for appreciative inquiry or sensemaking – can ensure experiences from the ground provide evidence of how outcomes work in context for specific people. These different perspectives also bring additional triangulation opportunities to enhance reasoning and credibility.

The need for bricolage also creates the opportunity to think about the entire process, starting from the way in which context is understood, to the development of outcome pathways as a process of co-production with multiple stakeholders. The WorldFish case illustrates this clearly through its use of PAR to inform programme and evaluation design. The greater the engagement with evaluation stakeholders throughout the process, the greater the claim to responsiveness that can be made, and ultimately the more likely it is that findings will be used. Often, participation is greater at the outset, but especially when using multiple methods (or parts of methods) the analysis is left to the expert evaluator. We see great potential for participatory sensemaking and validation of findings, which some of the methods reviewed suggest is possible.

However, navigating power requires more than participatory practice and good intentions. Evaluators engaging in bricolage will also need another set of relational and political competencies (Apgar and Allen 2021). Understanding the political landscape can help evaluators to decide when to push for particular combinations, and when to deepen participatory intent or call for more rigorous testing of the strength of evidence, as well as how to communicate findings to particular stakeholders. Alongside political understanding, excellent facilitation skills, humility, honesty, and reflexivity are important when engaging in the relational processes of sensemaking and value judgement. Finally, perhaps the core competency that underpins quality in bricolage is being able to balance principles with pragmatism. There is no such thing as perfect bricolage, so agreeing what is 'good enough' requires clear principles to ensure compromise does not lead to substandard evaluation practice.

8 Conclusion

In this CDI Practice Paper we responded to the growing acceptance of multi-method, complexity-aware and learning-focused approaches to impact evaluation. Evaluators regularly make use of bricolage in practice. Yet, the craft of bricolage is rarely acknowledged or discussed. It has only recently gained attention in the evaluation field. We argued that bricolage has a crucial role to play in maximising the relative strengths and bolstering the weaknesses of different qualitative evaluation methods.

We offered a guiding framework for how to do qualitative bricolage practically. We provided criteria for assessing methodological rigour (reasoning, credibility, responsiveness, utilisation, and transferability), and proposed core functions across the evaluation cycle, from context analysis to evaluation use to help evaluators identify where there may be complementarities and substitutions between different parts of methods. We discussed two cases that demonstrate how formative and summative evaluations illustrate the benefits of several key functions of another evaluation method, but not the method in full. Deciding which parts of which methods to include or exclude is a deeply political process. Bricolage is also an art: navigating different forms of power is a critical skill for effective bricolage and fundamental to enhancing the rigour of the process.

References

Alvarez, S. et al. (2010) 'Participatory Impact Pathways Analysis: A Practical Method for Project Planning and Evaluation', Development in Practice 20.8: 946–58, DOI: 10.1080/09614524.2010.513723 (accessed 15 August 2022)

Apgar, M. and Allen, W. (2021) 'Section Introduction: Participatory Monitoring, Evaluation and Learning: Taking Stock and Breaking New Ground', in D. Burns, J. Howard and S.M. Ospina (eds), The SAGE Handbook of Participatory Research and Inquiry, London: Sage Publications

Apgar, M.; Hernandez, K. and Ton, G. (2020) Contribution Analysis for Adaptive Management, ODI Briefing Note, London: Overseas Development Institute (ODI)

Apgar, M. et al. (2017) 'Getting Beneath the Surface in Program Planning, Monitoring and Evaluation: Learning from Use of Participatory Action Research and Theory of Change in the CGIAR Research Program on Aquatic Agricultural Systems', Action Research 15.1: 15–34, DOI: 10.1177/1476750316673879 (accessed 15 August 2022)

Aston, T.; Roche, C.; Schaaf, M. and Cant, S. (2021) 'Monitoring and Evaluation for Thinking and Working Politically', Evaluation 28.1: 36–57, DOI: 10.1177/13563890211053028 (accessed 15 August 2022)

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 (accessed 15 August 2022)

Befani, B. (2016) Choosing Appropriate Evaluation Methods Tool, London: Bond (accessed 15 August 2022)

Burns, D.; Howard, J. and Ospina, S.M. (eds) (2021) The SAGE Handbook of Participatory Research and Inquiry, London: Sage Publications

Chambers, R. (2015) 'Inclusive Rigour for Complexity', Journal of Development Effectiveness 7.3: 327–35, DOI: 10.1080/19439342.2015.1068356 (accessed 15 August 2022)

Denzin, N.K. and Lincoln, Y.S. (eds) (1999) The SAGE Handbook of Qualitative Research, 3rd ed., Thousand Oaks CA: Sage Publications

Derbyshire, H.; Barr, J.; Fraser, S. and Mwamba, W. (2016) Moving Targets, Widening Nets: Monitoring Incremental and Adaptive Change in an Empowerment and Accountability Programme, DFID Nigeria Briefing Paper, London: Department for International Development

Douthwaite, B.; Apgar, J.M. and Crissman, C. (2014) 'Monitoring and Evaluation Strategy Brief', Program Brief AAS-2014-04, Penang: CGIAR Research Program on Aquatic Agricultural Systems (accessed 15 August 2022)

Douthwaite, B. et al. (2017) 'A New Professionalism for Agricultural Research for Development', International Journal of Agricultural Sustainability 15.3: 238–52, DOI: 10.1080/14735903.2017.1314754 (accessed 15 August 2022)

Douthwaite, B.; Mayne, J.; McDougall, C. and Paz-Ybarnegaray, R. (2017) 'Evaluating Complex Interventions: A Theory-Driven Realist-Informed Approach', Evaluation 23.3: 294–311, DOI: 10.1177/1356389017714382 (accessed 15 August 2022)

Estrella, M. *et al.* (eds) (2000) *Learning From Change: Issues and Experiences in Participatory Monitoring and Evaluation*, London: Intermediate Technology Publications

Goertz, G. and Mahoney, J. (eds) (2012) *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*, Princeton NJ: Princeton University Press

Greenhalgh, J. and Manzano, A. (2021) 'Understanding "Context" in Realist Evaluation and Synthesis', *International Journal of Social Research Methodology* 24.3: 1–13

Guijt, I. (2020) 'Accountability and Learning: Exploding the Myth of Incompatibility between Accountability and Learning', in A. Fowler and C. Malunga (eds), *NGO Management*, London: Routledge

Guijt, I. and Roche, C. (2014) '**Does Impact Evaluation in Development Matter? Well, It Depends What It's For!**', *European Journal of Development Research* 26.1: 46–54, DOI: 10.1057/ejdr.2013.40 (accessed 15 August 2022)

Hargreaves, M. (2021) 'Bricolage: A Pluralistic Approach to Evaluating Human Ecosystem Initiatives', *New Directions for Evaluation Special Issue: Systems and Complexity-Informed Evaluation: Insights from Practice* 170: 113–24

Heron, J. and Reason, P. (1997) '**A Participatory Inquiry Paradigm**', *Qualitative Inquiry* 3.3: 274–94, DOI: 10.1177/107780049700300302 (accessed 15 August 2022)

HM Treasury (2020) *Magenta Book 2020: Supplementary Guide: Handling Complexity in Policy Evaluation*, London: HM Treasury (accessed 15 August 2022)

Jimenez, E. *et al.* (2018) *Mixing and Matching: Using Qualitative Methods to Improve Quantitative Impact Evaluations (IEs) and Systematic Reviews (SRs) of Development Outcomes*, CEDIL Inception Paper 5, London: Centre of Excellence for Development Impact and Learning (accessed 15 August 2022)

Lévi-Strauss, C. (1968) *The Savage Mind*, Chicago IL: University of Chicago Press

Maxwell, J.A. (2012) *Qualitative Research Design: An Interactive Approach*, 3rd ed., Thousand Oaks CA: Sage Publications

Maxwell, J.A. (2004) 'Causal Explanation, Qualitative Research, and Scientific Inquiry in Education', *Educational Researcher* 33.2: 3–11

Patton, M.Q. (2019) *Blue Marble Evaluation: Premises and Principles*, New York NY: Guilford Press

Paz-Ybarnegaray, R. and Douthwaite, B. (2017) '**Outcome Evidencing: A Method for Enabling and Evaluating Program Intervention in Complex Systems**', *American Journal of Evaluation* 38.2: 275–93, DOI: 10.1177/1098214016676573 (accessed 15 August 2022)

Piron, L.H. and Waldock, S. (2021) *Getting into the Politics of Why (Some) Governance Programmes Work*, From Poverty to Power, blog, 19 November (accessed 15 March 2022)

Preskill, H. and Lynn, J. (2016) *Redefining Rigour: Describing Quality Evaluation in Complex Adaptive Settings*, FSG Reimagining Social Change, blog, 1 February (accessed 25 November 2020)

Roche, C. (2015) 'The Politics of Juggling Multiple Accountability Disorder', in R. Eyben, I. Guijt, C. Roche and C. Shutt (eds), *The Politics of Evidence and Results in International Development: Playing the Game to Change the Rules?*, Rugby: Practical Action

Rogers, P.J. and Weiss, C.H. (2007) '**Theory Based Evaluation: Reflections Ten Years On: Theory-Based Evaluation: Past, Present, and Future**', *New Directions for Evaluation* 2007.114: 63–81, DOI: 10.1002/ev.225 (accessed 15 August 2022)

Stern, E. *et al.* (2012) *Broadening the Range of Designs and Methods for Impact Evaluations: Report of a Study Commissioned by the Department for International Development*, Working Paper 38, London: Department for International Development (accessed 15 August 2022)

Ton, G. (2021) 'Development Policy and Impact Evaluation: Learning and Accountability in Private Sector Development', in H. Zafarullah and A.S. Huque (eds), *Handbook of Development Policy*, Cheltenham: Edward Elgar Publishing

Ton, G. (2012) '**The Mixing of Methods: A Three-Step Process for Improving Rigour in Impact Evaluations**', *Evaluation* 18.1: 5–25, DOI: 10.1177/1356389011431506 (accessed 7 September 2022)

Vaessen, J.; Lemire, S. and Befani, B. (2020) *Evaluation of International Development Interventions: An Overview of Approaches and Methods*, Washington DC: International Bank for Reconstruction and Development/World Bank (accessed 15 August 2022)

Annexe 1

Methods reviewed and discussed

The 33 methods reviewed were: Realist Evaluation (RE); Contribution Analysis (CA); Process Tracing (PT); PT with Bayesian Updating; Contribution Rubrics; Outcome Mapping (OM); Rapid Outcome Mapping Approach (ROMA); Most Significant Change (MSC); Outcome Harvesting (OH); Ripple Effects Mapping (REM); Reality Check Approach (RCA); Qualitative Comparative Analysis (QCA); Actor Based Change; Qualitative Impact Assessment Protocol; Collaborative Outcomes Reporting (COR); General Elimination Methodology; Significant Policy Improvement; Multiple Lines and Levels of Evidence; Participatory Impact Assessment and Learning Approach; Strategic Action Research; SenseMaker; Causal Link Monitoring; Success Case Method; Strategy Testing; Developmental Evaluation; Causal Loop Diagrams; Strategic Assumption Surfacing and Testing; Process Monitoring of Impacts; Social Network Analysis; System Mapping and Dynamics; Scenario Planning; SearchFrames; and the What Else Test.

For further information, Vaessen *et al.* (2020) and HM Treasury (2020) offer guidance and critical appraisal of: RE; CA; PT; PT with Bayesian Updating; OM; MSC; and QCA. Aston *et al.* (2021) provide examples of PT, RE, and OH against the rigour criteria. Apgar and Allen (2021) also offer a discussion of MSC, OM, REM, SenseMaker, and the RCA.

Table A1 Key methods discussed in bricolage examples

Method	Brief explanation
Realist Evaluation	A form of theory-based evaluation that develops a programme theory including mechanisms, the context in which they might operate, and the outcomes. It asks what works, for whom, in what respects, to what extent, in what contexts, and how?
Contribution Analysis	A theory-based evaluation approach that iteratively maps available evidence against a theory of change, then identifies and addresses challenges to causal inference in six systematic steps.
Process Tracing	A theory-based method that develops mechanisms to explain how and why a cause (or set of causes) led to a particular outcome, and employs evidence tests.
Outcome Harvesting	Retrospectively identifies intended and unintended, positive and negative outcomes by collecting statements and evidence of what has changed, then, working backwards, determining whether and how an intervention has contributed to these changes.
Most Significant Change	A method that generates and analyses personal (often community) accounts of change and involves participants in deciding which of these accounts is the most significant – and why.
Contribution Rubrics	A hybrid of Process Tracing and Outcome Harvesting, which employs rubrics to assess levels of outcome, contribution, and strength of evidence.
Participatory Action Research	An orientation to knowledge generation that includes participants' experiential knowledge. As a collaborative methodology in the context of evaluation, it facilitates inclusion of different participant experiences and knowledge in making sense of how change could or has happened. It is often combined with iterative use of theory of change.
Participatory Impact Pathway Analysis	Engages participants in predicting how project outcomes can lead to social, economic, and environmental impacts by creating impact pathways through a participatory workshop, paying attention to social networks. It employs outcome logic models to support evaluation of the process of change.

> " 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. "

Centre for Development Impact (CDI)

The Centre is a collaboration between **IDS**, **Itad** and the **University of East Anglia**.

The Centre aims to contribute to innovation and excellence in the areas of impact assessment, evaluation and learning in development. The Centre's work is presently focused on:

(1) Exploring a broader range of evaluation designs and methods, and approaches to causal inference.
(2) Designing appropriate ways to assess the impact of complex interventions in challenging contexts.
(3) Better understanding the political dynamics and other factors in the evaluation process, including the use of evaluation evidence.

This CDI Practice Paper was written by **Tom Aston** and **Marina Apgar**.

The opinions expressed are those of the authors and do not necessarily reflect the views of IDS or any of the institutions involved.