Mixed-methods impact evaluation in international development practice: distinguishing between *quant-led* and *qual-led* models

James Copestake

Published online: 09 May 2024.

Submit your article to this journal

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

Mixed-methods impact evaluation in international development practice: distinguishing between *quant-led* and *qual-led* models

James Copestake 🅳

Department of Social and Policy Sciences, University of Bath Bath, UK

ABSTRACT

Despite being widely endorsed for more than two decades, the practice of mixed-methods impact evaluation (MMIE) remains confusing. This paper suggests that greater clarity can be achieved by distinguishing between *quant-led* and *qual-led* models of MMIE. The quant-led model gives most weight to *variance-based* epistemological approaches to causal attribution but can also incorporate *process-theory* approaches. The qual-led model relies mainly on a process-theory approach but incorporates quantitative data collection and analysis. After setting out the context, the paper sets out these conceptual distinctions. It then presents an illustrative case study of how the Qualitative Impact Protocol (QuIP) has been utilised within the two models. Third, the paper explores divergent support for the two models. We conclude with reflections on how wider recognition of the distinction between them can improve evaluative practice by deepening our understanding of multiple options for the integration of qualitative and quantitative aspects of impact evaluation. While mainly intended to be of practical relevance to those planning, conducting, and reviewing MMIEs, the paper is also relevant to wider concerns over the political economy of knowledge production and distribution.

Introduction

Mixed-methods impact evaluation (MMIE) is a route to identifying planned and unplanned outcomes of interventions, causal mechanisms underlying these effects, and the conditions under which these arise to assist both organisational learning and political accountability (Bamberger, Rao, and Woolcock 2010). The general case for combining quantitative and qualitative methods broadly rests on two arguments – that the strengths of each can mitigate the weaknesses of the other, and that their integration can add to the overall credibility of findings (e.g. Woolcock 2019, 4). Confusion persists over how to realise these potential payoffs in practice, MMIE being widely viewed as a worthy aspiration, but one that is difficult to do well (Bamberger 2015; Jimenez et al. 2018; Kabeer 2019; H. White 2011; S. White 2015). This paper argues that practice can be improved by enriching our understanding of the possibilities of qual–quant integration, including through wider recognition of the prevalence of two distinct (*quant-led* and *qual-led*) models of MMIE.

Impact evaluation can be distinguished from other forms of research by its practical focus on identifying outcomes of a specific intervention, whether a time-bound project or experiment, or a more open-ended programme or policy. This paper uses the distinction between *variance* and

process theory-based strategies for causally attributing outcomes to interventions to contrast two models of MMIE in international development practice. The first is labelled *quant-led* and relies mainly on variance-based causal attribution, but its use also entails doing qualitative tasks. It also accommodates process theory-based causal attribution in a complementary and subordinate way. The second is labelled *qual-led* and relies mainly on process theory-based attribution, but also incorporates collection and analysis of quantitative data. For example, realist evaluations often use quantitative data to identify variation in outcomes and contexts of interventions but rely mainly on process theory-based attribution to identify causal mechanisms (Pawson 2013).

Methodologically, the paper employs what critical realists call 'retroduction' (Bhaskar 2016). This proceeds in two steps. The first 'abductive' step is to formulate a probable causal explanation for an outcome based on prior knowledge. The second is to use a combination of deduction and induction to test this causal claim. In this paper, the abductive step is the proposition that confusion over MMIE in international development (an outcome) is partly attributable to the existence of two distinct models of MMIE.[1] This claim draws on prior knowledge derived from a combination of secondary literature review and the direct experience of using the Qualitative Impact Protocol (QuIP) summarised in the third section of the paper.[2] The second step is to deduce that some causal process must have led to the emergence of these two models of MMIE, and to seek a historical explanation for this – a task attempted in the fourth section of the paper.

The main purpose of the paper is to be of practical use to those planning, conducting, and reviewing MMIEs by aiding understanding of the range of approaches available. At the same time, it is relevant to the broader issue of the political economy of knowledge production and distribution, including the preference constraints and interests of impact evaluation commissioners and researchers. The quant-led model fits with a more positivist approach to social science, and a world view that favours relatively universal, technical, and linear forms of development intervention. In contrast, the qual-led model reflects an interpretive view of development that is more path-dependent, social, and complex. Application of the two models generate different kinds of evidence to serve distinct interests, and the paper does not claim or conclude that one is universally more useful in bringing evidence to bear on complex development issues than the other. This uncertainty warrants further research, as do questions about whose interests are best served by differences in current norms for controlling access to the findings arising from different approaches to MMIE.

Conceptual framing

An excursion into concepts and definitions can be justified by ambiguity and confusion over use of the core concepts of 'quantitative' and 'qualitative' that underpin the idea of mixed methods.[3] This section first distinguishes between two approaches to causal attribution – the challenge that lies at the heart of impact evaluation. It then develops a broader framework for thinking in a more granular way about quantitative and qualitative aspects of different approaches to MMIE. In so doing, the discussion negotiates three quite different ways of thinking about the qual–quant dichotomy – as two distinct cultures, as two sets of research methods, and as two poles on a spectrum of different kinds of research activity.

The focus of this paper is on collecting and interpreting evidence of how a specified intervention (X) has causal effects (Y), where the causal relationship between them is complicated by the presence of additional confounding factors (Z). Bold type indicates that X, Y and Z are vectors of factors. This nomenclature can be used to draw the distinction (taken from Maxwell 2004) between *variance based* (primarily quantitative) and *process theory based* (primarily qualitative) approaches to causal attribution. For researchers working with the variance-based approach, this entails identifying a counterfactual – i.e. what *would* have happened to Y if X had not happened, with Z remaining the same (Dunning 2012; Glennerster and Takavarasha 2013; H. White and Raitzer 2017). Only by comparing observed changes in Y with such a counterfactual is it possible to arrive at an internally valid measure of the causal effect of X. Since the counterfactual is unobservable such attribution

entails exploiting measurable variation in exposure to **X** across a large enough sample of cases to permit statistical estimation of the effect of **X** on **Y** while minimising the confounding effects of observable variation in **Z**. Randomised controlled trials (RCTs), natural experiments and other quasi-experimental methods offer a range of solutions to this problem (H. White and Raitzer 2017).

For those employing the process theory approach to attribution, latent counterfactuals or 'what if' scenarios also reside inside different stakeholders' heads, embedded in the language they use to explain *how* change happens.[4] An advantage of relying on what people say is that a self-contained set of claims about the causal mechanisms linking **X**, **Z** and **Y** can be collected from each independent source, revealing how the same intervention can have highly heterogeneous effects. However, such claims are susceptible to numerous forms of bias, arising both from deep cognitive processes and from how social positioning influences what we feel, think, and say (Hewstone 1989; H. White and Phillips 2012). Narrative evidence of causal claims is also conceptually fuzzy, and generalising from it is messy because it is not collected to fit a predetermined conceptual framework or coding pattern (Powell, Copestake, and Remnant 2024). The challenge facing researchers is to elicit and combine multiple sources of understanding in a way that can be subjected to critical scrutiny and contribute to useful generalisation. Contribution analysis, process tracing, realist evaluation and many qualitative impact evaluation methods partly address this problem by tailoring data collection and analysis to test one or more prior theories in a relatively transparent way (Stern et al. 2012; H. White and Phillips 2012). Findings are generalisable to the extent that they help to refine prior theories to explain the causal processes through which different combinations of **X** and **Z** lead to **Y**. This entails 'tangling' and synthesising multiple sources of evidence and theory logically together into 'middle-range theories' that usefully fill the vast chasm between universal laws, and causal explanations or 'theories of change' unique to one time and place (Cartwright 2020).

A 'belt and braces' or 'Q-squared' approach to MMIE makes parallel use of both these approaches to causal attribution in parallel, with interaction between them confined mainly to initial planning and final data interpretation stages. This is perhaps how many researchers think about MMIE, with the variance-based approach classified as a quantitative method, and the process theory approach as a qualitative method. This conflation of quant and qual with attribution methods is also consistent with the tendency for social scientists to specialise in using one or other approach, and even to associate such specialisation with different disciplines (Repko and Szostak 2021).[5] Distinguishing between relatively self-contained quantitative and qualitative methods also reflects a wider tendency to do so within disciplines. In political science, for example, Goertz and Mahoney (2012, 2) distinguish between quantitative and qualitative research traditions according to ' . . . whether one mainly uses within-case analysis to make inferences about individual cases (qual) or . . . cross-case analysis to make inferences about populations (quant)'.

Morgan (2007) acknowledges the power of this dichotomous approach by defining qualitative research as a primarily subjective process that is inductive in the way it links theory and data to draw context-specific inferences, contrasting this with quantitative research that is mostly deductive and aspires to make objective and generalised inferences. However, he also argues that a paradigmatic shift to methodological pluralism has eroded these distinctions: integrating induction and deduction through retroductive reasoning; emphasising intersubjectivity over the dichotomy between subjective and objective; and seeking cross-context transferability of causal theories without aspiring to establishing universal laws. This suggests scope for more nuanced thinking. For example, Haig and Evers (2016, 89) suggest that ' . . . in many cases, we will likely gain a better understanding of individual research methods we use, not by viewing them as either qualitative or quantitative in nature, but by regarding them as having both dimensions'.

A more granular way to differentiate between discrete quantitative and qualitative research tasks is to consider the extent and timing of data *codification*. More quantitative approaches code data *early* and in greater detail to facilitate their efficient collection, storage, and numerical analysis. In contrast, more qualitative approaches *delay* codification. This makes it harder to handle large amounts of data numerically but requires fewer assumptions about what data to 'admit' and in

Table 1. Possible quant-qual causal interactions within MMIE.

Phase	Quant-qual interactions *within* phases	Quant-qual interactions *between* phases
t = 1	quant1 -> qual1	quant1 -> qual2
Before intervention	qual1 -> quant1	qual1 -> quant2
t = 2	quant2 -> qual2	quant1 -> qual3
During intervention	qual2 -> quant2	qual1 -> quant3
t = 3	quant3 -> qual3	quant2 -> qual3
After intervention	qual3 -> quant3	qual2 -> quant 3

what form, thereby increasing the range of possible findings (Moris and Copestake 1993). The distinction can also be applied *within* methods, thereby permitting more granular descriptions of mixed-methods research designs. For example, the QuIP (see below) collects qualitative narrative evidence, incorporates the quantitative step of counting the frequency of causal claims coded across it, then uses this data to inform qualitative judgements about how far the evidence confirms a prior theory.

For impact evaluation purposes, an additional source of complexity is often the need to synchronise these tasks with the intervention being evaluated – by distinguishing between tasks carried out before (*t* = 1), during (*t* = 2) and after (*t* = 3) phases of the intervention, for example.[6] Adapting the nomenclature favoured by Creswell and Plano Clark (2018), interactions between quantitative and qualitative tasks can be identified both *within* each phase, and *between* them (i.e. from one phase to a later phase), as shown in Table 1. The table distinguishes between 12 different categories of qual–quant interaction, and hence suggests a huge number of designs of MMIE are possible, given that one method may incorporate several of them. Additional variation also arises from the relative *weight* and *purpose* of each quant and qual component (Guest and Fleming 2015).

A further elaboration of this framework would distinguish between three methodological tasks: (a) framing and planning, (b) data collection, and (c) data analysis and use. Framing and planning *may* mostly take place before the intervention (*t* = 1), data collection during its implementation (*t* = 2), and analysis after it finishes (*t* = 3), but it is rarely this simple. For example, designs for evaluation of adaptive management projects may need to be adjusted after they start; difference-in-difference studies often include pre-intervention data collection in the form of a baseline survey so as to reduce reliance on respondent recall; and natural experiments can be conceived after an intervention is over (Dunning 2012), whereas planning an RCTs entails allocating treatment of cases before the intervention starts, and so on.

Having set up a framework for examining many possible designs of MMIE the remainder of this section uses it to distinguish between two leading models of MMIE, referred to as *quant-led* and *qual-led*. These differ primarily according to the weight ascribed to primarily quantitative (variance based) and qualitative (process theory based) causal attribution. The distinction is an abductive leap, which draws on a wide but unsystematic review of published literature on MMIE along with direct experience of using the QuIP (see below). As such, it is provisional, being primarily intended to aid understanding of the unsettled nature of current MMIE practice in international development practice, as also discussed later in the paper.

The quant-led MMIE model

At its simplest, this can be broken down into three phaseses: integrated design, parallel data collection and analysis, and integrated interpretation. This can be expressed as follows:

(QUANT1 ⟵→ qual1) →
(qual2, quant1&3) →
(qual3⟵→QUANT3)

where the double arrow indicates two-way interactions, and capitals indicate subordination of one component to another.

My depiction of this model primarily generalises from a systematic review of 40 impact evaluation studies conducted by Jimenez et al. (2018).[7] I then checked it against more recently published examples (de Allegri et al. 2020; de Milliano et al. 2021; Margolies et al. 2023, and Ranganathan et al. 2022), as well as more critical perspectives on quant-led MMIE by White (2015) and Kabeer (2019). Table 2 elaborates on key tasks and their interactions across the three intervention phases.

In the first phase, the main qualitative task is to inform design of the variance-based impact evaluation. This includes contextual analysis, appraising the efficacy, acceptability and ethics of conducting an evaluation, refining the causal theory of change informing evaluation design, identifying key concepts to include, developing measurable indicators for them, and pilot testing research instruments (Garcia and Zazueta 2015; H. White 2011). Once key research questions are agreed, then statistical power calculations play an important role in determining minimum sample sizes needed to produce statistically significant results, and hence the cost of data collection. The methodology for determining how large any parallel process theory-based impact evaluation should be is less precise. It also hinges on using available data (including from any baseline survey) to ensure qualitative case and source selection picks up as much of heterogeneity in the intervention's impact as possible – a key quantitative input into process-theory-based impact evaluation (Copestake 2021). This methodological difference can have an important bearing on the relative allocation of funds between the parallel impact evaluation efforts, reinforced by differences in expectations about what each will deliver.

Typically, a baseline survey provides the foundation for quantitative impact assessment, followed by at least one post-intervention or so-called endline survey, permitting statistical analysis of correlations between observed changes in **Y** across the sample and variable exposure to **X**, while

Table 2. The quant-led MMIE model.

	Intervention phase		
	Before ($t = 1$)	During ($t = 2$)	After ($t = 3$)
Main evaluative activities	Needs assessment. Scoping and framing. Design and planning. Appraisal of the intervention. Evaluability assessment.	Process evaluation. Timely learning. Mid-course adjustments.	Reflect on what to do next: e.g. continue, scale-up, adjust or terminate. Account to stakeholders.
Main qualitative components	Scoping and framing. Stakeholder consultation. Conceptualisation and indicator selection. Formulating 'espoused' theories of change.	Mid-term qualitative interviews. Review 'in-use' theories of change.	Interpret observed outcomes. Review and revise 'espoused theories' of change.
Main quantitative components	Scoping surveys. Baseline survey. Power calculations.	Mid-line survey (optional).	End-line survey(s). Statistical analysis of survey data. *Ex-post* cost-benefit analysis.
Qual-> Quant interactions	Extensive qual inputs into design of quant tasks at all stages.	Limited because quant design is now set.	Qual findings can aid interpretation of causal processes underpinning quant findings and their generalisability.
Quant-> Qual interactions	Quant data can usefully inform source and case selection for the qual components.	Quant data can inform judgements about the context and generalisability of the qual findings.	Quant data can inform judgements about the context and generalisability of the qual findings.
Power and resource issues	Quant components dominate the budget and are more time critical. Qual specialists are often marginalised.	Qual components are often sub-contracted, and findings marginalised.	Quant findings focused on impact eclipse qual findings centred on relevance.

controlling also for variation in **Z**.[8] Qualitative data collection and analysis proceeds in parallel and is used to collect more detailed evidence of the causal mechanisms linking the intervention, contextual factors and specified outcomes, typically relying mostly on narrative accounts of the processes collected through interviews, focus groups, and other relevant written material.[9] An optional extra is for this to continue into the post-implementation phase, including tailored research into unanswered questions thrown up by the quantitative impact evaluation – see Gibbs et al. (2020), for example.

In the third stage of the evaluation, findings obtained in the parallel qualitative and quantitative strands are compared, with a particular emphasis on how far causal pathways and mechanisms identified qualitatively can help to explain statistically significant correlations between **X, Y** and **Z**. In addition, the qualitative evidence can be used to throw light on reasons for variation in impact between different individuals and groups within the selected population, given limitations in the extent to which the quantitative analysis can go beyond evidencing *average* 'intent-to-treat' or 'treatment-on-the-treated' effects.

Recent published examples suggest that the broad pattern of quant-led MMIE is relatively settled: the main differences lying in detailed design of the two strands, and how fully and effectively the qualitative component is integrated into interpretation of the quantitative findings. More minimalist studies relegate qualitative methods to mapping the context and assessing implementation fidelity, rather than contributing to causal inference. More comprehensive studies pay closer attention to how a process-theory-based strand maps onto the quantitative dataset to support credible inferences about the operation of context-specific causal mechanisms. For an example, Bonilla et al. (2017) draw on a qualitative strand to identify causal mechanisms consistent with quantitative findings, avenues for quantitative analysis of heterogeneous impact, and scope for improving the robustness of selected indicators of women's empowerment.

Despite the potential for integration of the two approaches to causal attribution, there is a strong tendency for the variance-based strand to dominate. A key explanation for this is its promise to meet commissioners' demands for defensible and precise answers to cost–benefit questions. White (2015) also highlights insufficient involvement of experienced qualitative researchers in the design and management of such studies.

The qual-led MMIE model

This model centres on qualitative assessment of impact supported by quantitative monitoring. Like the quant-led model, it can also be set out across the three phases of a project intervention, but in its purest and simplest form it looks like this:

(quant2←→QUAL2)

This is because the model is particularly suited to evaluation of open-ended programmes, and policies, as well as the strategies of organisations based on a rolling portfolio of time-bound projects.

Initial qualitative activities include consulting stakeholders, clarifying key concepts, and making explicit the theory of change underpinning the intervention. These in turn inform development or modification of an information system for real-time monitoring of key indicators of **X**, **Y** and **Z** at different levels of aggregation. Such systems are mostly designed to support routine performance management, enabling internal and external stakeholders to make their own judgements about possible causal processes explaining expected and unexpected patterns and variations in the data over time. They also provide a strong quantitative foundation for supplementary impact evaluation to verify or challenge these judgements. Larger organisations institutionalise choices about how much, when, and why additional evidence is needed through the employment of specialised monitoring and evaluation staff. It is also consistent with reliance on external evaluation studies to supplement internal evaluation and decision support systems. The model resonates with both an opportunistic or *bricolage* approach to MMIE (Aston and Apgar 2022; Heinemann, Van Hemelrijck,

Table 3. The qual-led impact evaluation model.

Phase	During ($t = 2$)
Main evaluative activities	Process evaluation.
	Timely learning.
	Adaptive management.
Mainly qualitative components	Data collection bricolage.
	Review cycles.
	Feedback on prior theories of change.
	Incremental adjustment of the impact evaluation approach.
Mainly quantitative component	Continuous monitoring of key performance indicators at multiple levels.
	Variance based impact evaluation may also be incorporated.
Qual-> Quant interactions	Qual findings inform continuous adjustments to quant monitoring systems.
	Qual findings can also inform quantitative modelling and simulation.
Quant-> Qual interactions	Quant data informs judgements about the context and generalisability of unfolding qual findings.
Power and resource issues	Impact evaluation supplements ongoing performance management.
	Evaluative activities are responsive to perceived management needs.
	Integrated analysis of evidence is often hidden from external stakeholders.
	Risks of partial and distorted understanding of impact arising from overemphasis on any one measure or approach.

and Guijt 2017), and with a more formal Bayesian approach (Humphreys and Jacobs 2015). Both process and variance-based impact evaluation may feature as part of the mix of methods used to challenge and refine an organisation's ongoing understanding of its impact.[10]

While consistent with a 'complexity-informed' view of development practice (Chambers 2015; Hernandez, Ramalingam, and Wild 2019; Rogers 2020) this approach is also an extension of routine performance monitoring and management. For example, fire safety systems for buildings build on continuous quantitative monitoring using smoke detectors to provide binary data on the presence or absence of smoke in multiple locations at any moment. But they also depend on timely qualitative feedback to explain why alarms are triggered or failed – all informed by strong underlying theory about the causes and consequences of fire.[11] Two-way qual–quant interactions are critical to the model, with qualitative data collection and interpretation informing choice of key monitoring indicators, as well as how, how frequently, and at what level of aggregation they are collected, analysed and shared. In the reverse direction, identification of trends and other patterns from monitoring quantitative indicators informs specification and focus of qualitative impact evaluation – see Table 3.

Illustrative case study: use of the QuIP in MMIE

This section provides a case study of how a process theory-based impact evaluation approach – the Qualitative Impact Protocol (QuIP) – has been incorporated into different MMIE designs. It briefly describes the QuIP then reviews how 64 QuIP studies listed in the Appendix contributed to MMIE studies based on both the quant-led and qual-led models.

The QuIP was designed through collaborative action research led by the University of Bath, and mainstreamed by Bath SDR Ltd, a social enterprise set up specifically to broaden the range of approaches to impact evaluation (Copestake, Morsink, and Remnant 2019, 6). It relies on collecting narrative causal statements directly from those affected by an intervention using a mix of semi-structured interviews and focus group discussions. By framing these around experience of change in selected outcome domains, and exploring perceived reasons for these, equal weight is given to all possible explanations for the changes identified. Where possible, data collection is also 'double blindfolded', meaning that field investigators and respondents are provided with as little information as possible about the specific intervention being evaluated.[12] Another feature of the QuIP is that rather than relying on analysis of text through thematic coding of *concepts* it relies on directly coded *causal claims* embedded in narrative text – each claim linking at least one causal driver and one outcome or effect. This facilitates analysis and interpretation of

findings visually using causal maps (Copestake, Davies, and Remnant 2019; Powell, Copestake, and Remnant 2024).

The QuIP is particularly intended to enhance understanding of complex situations – revealing unexpected causal factors and unintended outcomes, delving deeply into the causal mechanisms at play, and confirming or challenging prior theories of change (Copestake 2014). It does not set out to generate precise quantitative estimates of causal connections, such as average treatment effects, nor data that are statistically representative of the views of a wider population of stakeholders. For these reasons, it is particularly suited to being used alongside quant methods.

Detailed guidelines for a QuIP are based on a benchmark study comprising 24 interviews and four focus group discussions, with the precise number and selection of respondents adjusted to suit the context and needs of a particular study. Purposive and stratified selection of sources and cases is employed to increase the probability of picking up as much diversity of experience as possible, including anomalous, positive, and negative deviant cases. This is aided by being able to draw on previously collected quantitative data on **X**, **Y** and **Z**. Source selection for confirmatory analysis is further strengthened by drawing upon the theory of change underpinning the intervention (Copestake 2021).

The Appendix lists all 64 discrete impact evaluation studies using the QuIP in which Bath SDR directly participated between 2016 and 2023. These studies were conducted across 24 countries for 28 different organisations, including local and national government agencies, charities, foundations, private companies, impact investors and bilateral donors. Twenty-five were primarily focused on promoting agricultural and rural development, 17 on health promotion including training health workers, and 22 on a wide range of other activities. The studies assessed how far selected projects and programmes were delivering intended benefits to defined target groups who included farming households, factory employees, students, users of public services, users of financial services, small/micro business owners, community-level organisations and NGOs. The mean number of interviews per study was 38, supplemented by 4.5 focus groups.

Only eight of the 64 listed studies were explicitly connected with a variance-based impact evaluation, and the strength of qual–quant interaction between them varied widely. The example that most closely conforms to the quant-led model comprised three rounds of QuIP studies alongside a randomised controlled trial (RCT) of a pilot poverty graduation programme in Malawi (Concern Worldwide 2021). This was conducted jointly with Trinity College Dublin, with survey data for the RCT used to inform selection of sources for the QuIP studies, and findings from the QuIP studies used to inform discussion of possible causal mechanisms explaining findings of the RCT. Similar, but more limited two-way interactions took place between QuIP and 'difference-in-difference' studies with Oxfam UK in Ethiopia, PDA Associates in Ghana, and ITAD Ltd in Nepal (Hedley and Freer, 2022); whereas there was no interaction at all between a QuIP and a parallel difference-in-difference study for the C&A Foundation in Mexico (Copestake, Morsink, and Remnant 2019, 75). Lastly, in Tanzania, the QuIP study drew on baseline data from an RCT, but only after the RCT itself was abandoned (Copestake, Morsink, and Remnant 2019, 142). The practical difficulties of realising the full potential of MMIE is illustrated by the failure – even once – to be able select respondents for a QuIP purposively using measured changes in outcome indicators derived from prior longitudinal surveys.[13]

In all other instances listed in the Appendix self-contained QuIP studies contributed to ongoing qualitative assessment of an activity in a way more consistent with the qual-led model of MMIE. Selection of sources and cases for QuIP studies nearly always drew upon quantitative baseline or operational data of some kind, such as lists of housing loan recipients under the Habitat for Humanity study in India (Copestake, Morsink, and Remnant 2019, 95). Otherwise, the design and implementation of QuIP studies was not closely integrated with other impact evaluation studies. However, QuIP studies were mostly used to inform assessment of ongoing programmes and rolling portfolios of projects, subject to ongoing monitoring and evaluation activities of internal staff and

hired researchers. Tearfund, for example, commissioned a sequence of four QuIP studies of its multi-country Church and Community Mobilisation programme alongside but independently of a large ongoing survey, while Save the Children conducted three QuIP studies of a family of integrated agriculture, nutrition, WASH and childcare projects (Copestake, Morsink, and Remnant 2019, 117 & 141). In contrast, use of QuIP by other organisations appeared – from an outsider's perspective – to be more *ad hoc*. Edufinance, for example, runs a large global programme, and commissioned one QuIP study specifically to investigate how its work in Kenya was being affected by Covid-19, as one of range of stand-alone impact evaluation studies.

Explaining the coexistence of two models of MMIE

Having presented a framework that is consistent with many forms of MMIE, and illustrated this with reference to how the QuIP has been utilised within different approaches, this section returns to the core argument of the paper that it is useful to distinguish between quant-led and qual-led models of MMIE to aid understanding of contemporary impact evaluation practice and scope for its improvement. To support this argument, I move from normative discussion of the distinction between them to a historical institutional analysis of how they relate to the evolution of evaluative thinking and practice.

Of the two, it is the quant-led model of MMIE that has been more prominent in recent academic and policy debates over impact evaluation in the field of international development. One explanation for this is its association with the growth in micro-level public health, education, livelihood promotion and social development projects intended to nudge intended beneficiaries into changes in their knowledge, attitude, and behaviour (Banerjee and Duflo 2012). The growth of such projects also reflects the advantages to international donors of relatively technocratic interventions with measurable impact goals that can be replicated and scaled-up across diverse contexts, using results-based project management and what Schwandt and Gates (2021) describe as 'conventional' models of evaluation. The potential to achieve scale across large populations also justifies relatively lumpy investment in impact evaluation, using 'large-n' variance-based methodologies capable of delivering precise and relatively easy to interpret estimates of impact on predetermined indicators easily linked to the SDGs.

Alongside this kind of intervention are more flexible modalities of development practice based on the idea of 'adaptive management' that aim to address more complex political, institutional, and structural development problems (Andrews, Pritchett, and Woolcock 2012; Boulton, Allen, and Bowman 2015; Ramalingam 2013). These have been associated with support for evaluative practices better attuned to uncertain impact trajectories (Woolcock 2009), identification of unintended consequences (Bamberger, Tarsilla, and Hesse-Biber 2016), and to informing timely programme adjustments (Webster et al. 2018). Appreciation of contextual and operational complexity also explains increased interest among development professionals in alternative approaches to impact evaluation (Brousselle and Buregeya 2018; Stern et al. 2012; H. White and Phillips 2012). The qual-led model of MMIE is more congruent with this second strand of development practice, and with models of evaluative practice described by Schwandt and Gates (2021) as 'expanded conventional' and 'emerging alternative'.

It is possible to imagine a world in which development professionals select different policies, programmes, and projects to address higher-level goals, and then select an appropriate impact evaluation approach to fit. However, evaluation is affected not only by objectively task-specific 'best practices' but also by professional commissioners' wider interests and preferences (Martens et al. 2002). For example, their interest in impact evaluation – or lack of it – depends on how much importance they attach to empirical evidence at all compared to the political 'warm glow' of being seen to do good works (Copestake, O'Riordan, and Telford 2016). Their methodological choices are also be limited by 'preference constraints' and limited 'navigational capacity' arising from their own technical training in research methods (Rao and Walton 2004), and by dominant disciplinary norms.

For example, a strong commitment to quant-led approaches to impact evaluation may reflect commissioners' unfamiliarity with realist and complexity-informed understanding of the social sciences (Bhaskar 2016; Boulton, Allen, and Bowman 2015).

Demand for evidence is also influenced by the interests and preferences of evaluation specialists and researchers about how to supply it (Dahler-Larsen 2011; Eyben 2013; Hayman et al. 2016). A leading example in the field of international development is the well-documented 20-year growth of donor investment in RCTs after 2003 (Bédécarrats, Guérin, and Roubaud 2020; Camfield and Duvendack 2014; Howard 2022; Kinstler 2024; H. White 2019). This can be attributed in part to their fit with the technocratic genre of development projects described above. Advocates of RCTs also narrowed methodological debate to focus away from questions of wider relevance (including external validity) and towards the theoretical internal validity of RCTs compared to other variance-based solutions to the attribution problem. They were also able to emphasise their ability to deliver relatively precise and easily interpreted estimates of average treatment effects to inform cost–benefit calculations and judgements. The critical pushback that RCTs attracted (Basu 2014; Deaton and Cartwright 2018; Ravallion 2018; Rodrik 2008) casts doubt on how far the power of these ideas alone sustained the RCT bubble; other possible explanations include its congruence with a wider 'evidence revolution' (H. White 2019), a simplistic view of the transferability of natural science empiricism to the social sciences, and powerful incentives to conducting RCTs as a route to academic success (Kinstler 2024). Either way, having persuaded many evaluation commissioners that RCTs amounted to a 'gold standard', proponents of them have been well placed to endorse a supporting role for process theory- based methods within quant-led approaches to MMIE.

Support for RCTs can usefully be contrasted with an alternative perspective on MMIE even more entrenched in international development practice. Molecke and Pinkse (2017) distinguish between four practical arguments for scepticism about *any* source of evidence of impact: key outcomes cannot be measured credibly, doing so is too expensive, insufficient data is available to support credible causal claims, and the causal claims that can be supported are irrelevant. Those holding such views may not reject impact evaluation entirely, nor efforts to improve it. But they are likely to be less inclined to support expensive independent quant-led MMIE, and to be more favourably disposed towards qual-led approaches that dovetail with their appreciation of the value of insider understanding. This view is corroborated by social scientists who emphasise complexity, and the role that experience-based wisdom or *phronesis* plays in interpreting multiple sources of evidence, including those based on personal observation and trusted relationships (Boulton, Allen, and Bowman 2015; Flyvbjerg 2006; Nicholls, Nicholls, and Paton 2015, 276; Pritchett, Samji, and Mammer 2013).[14]

Beyond personal philosophical positions, practitioners' distrust of formalised IE is also tangled up with experience of the administrative and political risks associated with it. To illustrate, take the case of a fictional development organisation – *ABC*. Confronted by a complex reality, *ABC* relies on a set of general 'theories of action' to inform its decisions, including (a) the 'espoused theory' set out in promotional material, policies, and procedures, (b) informal 'theories-in-use' embedded in routine practices and the 'shared mental models' of staff (Argyris and Schon 1978; Denzau and North 1994; Senge 1990). A central role of *ABC*'s leadership is to manage tensions arising from the tendency for theory espoused at the top of organisation to become decoupled from everyday practices and 'theories-in-use' lower down the organisation and across collaborating organisations (Boxenbaum and Jonsson 2017). In this context, formalised evaluation can be viewed as a form of political deliberation that not only has possible instrumental value (to facilitate learning, demonstrate goal achievement, account to stakeholders) but also potential for misuse within wider struggles over organisational reputations and legitimacy (Alkin and King 2016, 2017; Deephouse et al. 2017).

This brief excursion into organisational institutionalism illustrates why leaders of development agencies may be cautious about both commissioning impact evaluations and sharing the findings. If *ABC* invests in a mix of internal and external evaluative activities, the internal process of learning from them remains largely invisible to external stakeholders. Experience with conducting QuIP, for example, has often included being unable to assess how findings contributed to cumulative insider

understanding of the impact of the interventions we were studying. Evaluating the impact of any source of evidence on complicated management decisions is itself methodologically difficult, and so the reluctance of commissioners to reveal how they arrived at key decisions is understandable even if frustrating to interested external stakeholders. Of course, commissioners of RCTs and quant-led MMIE are also open to reputational damage if they accede from the outset to independent publication of the findings, but this may be a risk worth taking when linked to funding for scaled-up interventions and replications.

A consequence of their closer association with adaptive approaches to the management of development is that specialists in qual-led MMIE have often found it hard to secure the permission of pragmatic commissioners to publish findings, in sharp contrast to the stimulus to publication arising from quant-led MMIE's association with a more technical and projectised view of development, the RCT bubble and the wider 'evidence revolution' celebrated by White (2019). It may be a strength of qual-led MMIE that the gap between performance management and formal evaluation is less, but this proximity also seems to be associated with some loss of freedom to share findings with peers.

The difference in power to publish findings among MMIE providers probably also reflects greater agreement among quant-led providers over quality standards and benchmarks. In contrast, lack of publication weakens feedback loops through which standards for adaptive use of qual-led MMIE could be improved. Contributors to qual-led MMIE must also weigh up restrictions on how widely findings are shared with the benefits of building trust and some influence. However, polarisation of quant-led and qual-led approaches based on divergent transparency and researcher incentive structures contributes to general confusion about MMIE that helps nobody. In contrast, clearer understanding of the difference between them could foster wider recognition of the scope for strengthening integration of both process theory attribution within quant-led approaches and variance-based attribution within qual-led approaches – e.g. through realist RCTs.[15] There is also scope for further clarification of the similarities and differences between qualitative methods, and for building stronger standards for discrete qualitative impact evaluation studies (QuIP being just one example) to facilitate their wider publication, even, while they remain – and are understood to remain – only one component of the multi-strand MMIE that guides commissioning organisations.

Conclusions

Despite the existence of mixed-methods social research as a distinct field, widespread professional specialisation in quantitative or qualitative research methods persists and contributes to confusion over mixed-methods impact evaluation (MMIE) in international development practice. The aim of this paper is to go beyond this simplistic dichotomy. First, it offers a conceptual framework for more fine-grained analysis of the use of qualitative and quantitative methods to demonstrate that there are many possible forms of MMIE. Second, it suggests a useful distinction between just two – a quant-led and a qual-led model informed both by secondary literature, and direct experience of using the Qualitative Impact Protocol (QuIP) within different MMIE designs. The distinction between the two models is then explored further by relating it to recent trends and debates over impact evaluation in international development practice. This final section extends the argument, reflects on its limitations, and suggests possible directions for future research and practice.

The quant-led model of MMIE is centred on variance-based attribution, supported by qualitative contextualisation and design, and supplemented by process theory-based attribution to help explain findings. It fits with a more positivist approach to social science, and a relatively replicable, technical, and linear view of development practice informed by answers to relatively stable and narrowly defined causal questions. While costly to produce it has the potential to come up with relatively easily understood and scientifically credible numbers for the magnitude of development impact that commissioners demand, even while leaving open the question of how relevant these findings are to other contexts.

The qual-led model combines quantitative monitoring with reliance on process theory-based attribution, combining multiple sources of evidence in an open-ended 'Bayesian' process of testing and updating theoretical understanding of causal mechanisms. It reflects an interpretive view of development that is more path-dependent, social, and complex. Findings tend to be less precise but can be broader in scope, informing reflection over their relevance to other contexts, picking up on unexpected causes and effects, and enriching understanding of underlying causal mechanisms.

The reason for highlighting the existence of these two models is not to argue in favour of one over the other. Rather they serve as contrasting reference points and a counterpoint to the idea of there being a single 'best practice' model for MMIE. At the same time, the paper does also suggest multiple avenues for improving practice. First, quant-led MMIE can move closer to more equal 'belt-and-braces' integration of variance-based and process theory-based approaches within a single study. Second, scope remains for better and more consistent use of process theory-based approaches, on their own and as part of MMIE designs, and to a stronger commitment to allowing their publication on the part of those who commission them. Third, there is scope for wider recognition that variance-based and quant-led studies are always nested in wider, more complex, and qual-led processes of making judgements about 'what works', where, when, and for whom.

Moving beyond technical discussion of methodology, this paper also reflects on the political economy of impact evaluation, including the path-dependent preferences and interests of commissioners and researchers. More specifically, it suggests that asymmetry in norms affecting the publication of findings from quant-led and qual-led approaches are an obstacle to better understanding of the range of MMIE options, and to progress towards better practice. Growing political pressure to decolonise development practice should encourage more reflection on the highly unequal distribution of power and influence over how evidence of impact is conceptualised, produced, and distributed.

Notes

1. The term 'model' is used here as shorthand for what Denzau and North (1994, 3) refer to as the 'shared mental models' that in the presence of strong uncertainty 'guide choice and shape the evolution of political and economic systems and societies'.
2. The QuIP is an approach to impact evaluation based on collecting, coding, analysing and mapping narrative accounts of the causal drivers of change (Copestake 2021; Copestake et al. 2018, 2019a, 2019b).
3. The focus here is mainly on *mixed* methods ('qual-quant') rather than *multi*-methods ('quant-quant' and 'qual-qual') because it is widely viewed as more challenging (Fetters and Molina-Azorin 2017). Tashakkori and Creswell (2007) and Creswell and Plano Clark (2018) elaborate further on the definition of mixed method research in general.
4. The idea of latent counterfactuals builds on what Harari (2011) calls the 'cognitive revolution' through which the human species developed the capability to imagine other scenarios and thereby anticipate danger, plan, survive longer and sometimes even thrive. Reichardt (2022) offers a comprehensive review of different forms of counterfactual thinking.
5. An illustrative example of this is how Rao (2022, 5) identifies four ways in which quantitative economics can learn from the use of qualitative methods in social anthropology and sociology to become more reflexive – by developing cognitive empathy, learning to analyse narrative text, understanding processes of change, and using participatory methods to democratise otherwise ethically dubious processes of data extraction.
6. Of course, evaluations must often also synchronise with multi-stage interventions, including piloting and mainstreaming (Picciotto and Weaving 1994), unplanned interruptions and adjustments. Webster et al. (2018) explore the timing of impact evaluation in more detail.
7. The criteria they use for assessing the *mixed methods component* are specification of a clear theory of change, integration of methods at the design stage (including clarity about when and how qualitative evidence is to be used), integration of methods to inform interpretation of findings, and discussion of the limitations to integration. They conclude that the best studies provide a clear rationale for integration of methods, deploy multidisciplinary teams, adequately document what they do, and are open about the generalisability of findings.

8. Many studies also include a 'Quant2' or mid-line survey through which intermediate impact can be assessed quantitatively before the intervention ends.
9. Pierotti (in Goldstein and Pierotti 2020) draws on World Bank experience to emphasise the role of qualitative methods in understanding 'meaning and motivations', including the stories people tell themselves when they make decisions.
10. This menu of choices includes what White and Phillips (2012) call 'Group 1' approaches that explicitly set out to discover the causes of observed effects (realist evaluation, general elimination methodology, process tracing and contribution analysis), and Group 2 approaches that prioritise stakeholder participation (most significant change, the success case method, outcome mapping and MAPP). For discussion of these options see also Stern et al., (2012), Copestake et al. (2019b, ch.2), and Chambers (2009). Variance based approaches can also contribute to the flow of evidence, including approaches like interrupted time-series analysis and natural experiments that based on administrative data.
11. Gawande (2008) provides powerful insights into this way of thinking, while Eyben (2013) and Honig and Pritchett (2019) explore traps arising from over-reliance on mechanical use of quantitative targets to drive performance.
12. This design strategy aims to mitigate the risk of confirmation and strategic response biases – i.e. people saying what they think researchers want to hear, or will serve their own best interests. See Copestake et al. (2018) for a fuller discussion of this, and Copestake et al. (2019b) for discussion of how to mitigate bias arising from the positionality of researchers.
13. A partial exception to this were QuIP interviews conducted for the UK Home Office with Kantar Public. These were selected purposively from among respondents to an online survey, according to how they responded to questions about how much more or less secure they felt when walking in the streets at night.
14. The fundamental difference with those who advocate RCTs is ontological rather than epistemological: that empirical rigour is achieved only by making simplifying assumptions about the complexity of emergent social processes, hence the possibility of identifying regularities in causal relationships across different contexts.
15. For discussion of the scope for econometric analysis within realist research and evaluation see Olsen and Morgan (2005), Olsen (2019), Morgan (2019) and Warren et al. (2022).

Notes on contributor

James Copestake is Professor of International Development, in the Department of Social and Policy Sciences at the University of Bath, where he is also affiliated to the Centre for Development Studies, Centre for Qualitative Research, and Institute for Policy Research. His experience of research, doctoral supervision and publication span the following: agrarian change and rural development; development finance, microfinance and impact evaluation; definition and measurement of poverty and wellbeing; the political economy of international development. He is particularly interested in the interaction between them.

ORCID

James Copestake 🔵 http://orcid.org/0000-0001-5658-2597

References

Alkin, M. C., and J. A. King. 2016. "The Historical Development of Evaluation Use." *American Journal of Evaluation* 37 (4): 568–579. https://doi.org/10.1177/1098214016665164.

Alkin, M. C., and J. A. King. 2017. "Definitions of Evaluation Use and Misuse, Evaluation Influence and Factors Affecting Use." *American Journal of Evaluation* 38 (3): 434–450. https://doi.org/10.1177/1098214017717015.

Andrews, M., L. Pritchett, and M. Woolcock. 2012. "Escaping Capability Traps Through Problem-Driven Iterative Adaptation (PDIA)." *Working Paper 299*, Washington DC: Center for Global Development. Accessed August 8, 2022. https://www.cgdev.org/publication/escaping-capability-traps-through-problem-driven-iterative-adaptation-pdia-working-paper.

Argyris, C., and D. Schon. 1978. *Organizational Learning: A Theory of Action Perspective*. Reading, MA: Addison-Wesley.

Aston, T., and M. Apgar. 2022. "The Art and Craft of Bricolage in Evaluation." *Centre for Development Impact Practice Paper, No.24*, www.idea.acv.uk/cdi.

Bamberger, M. 2015. "Innovations in the Use of Mixed Methods in Real-World Evaluation." *Journal of Development Effectiveness* 7 (3): 317–326. https://doi.org/10.1080/19439342.2015.1068832.

Bamberger, M., V. Rao, and M. Woolcock. 2010. "Using Mixed Methods in Monitoring and Evaluation: Experiences from International Development." *Policy Research Working Paper, 5245*, World Bank.

Bamberger, M., M. Tarsilla, and S. Hesse-Biber. 2016. "Why so Many 'Rigorous' Evaluations Fail to Identify Unintended Consequences of Development Programmes: How Mixed Methods Can Contribute." *Evaluation and Program Planning* 55:155–162. https://doi.org/10.1016/j.evalprogplan.2016.01.001.

Banerjee, A. V., and E. Duflo. 2012. *Poor Economics: Barefoot Hedge-Fund Managers, DIY Doctors and the Surprising Truth About Life on Less Than $1 a Day*. London: Penguin Books.

Basu, K. 2014. "Randomisation, Causality and the Role of Reasoned Intuition." *Oxford Development Studies* 42 (4): 455–472. https://doi.org/10.1080/13600818.2014.961414.

Bédécarrats, F., I. Guérin, and F. Roubaud. 2020. *Randomized Control Trials in the Field of Development: A Critical Perspective*. Oxford: Oxford University Press. ISBN: 9780198865360.

Bhaskar, R. 2016. *Enlightened Common Sense: The Philosophy of Critical Realism*. London: Routledge.

Bonilla, J., R. C. Zarzur, S. Handa, C. Nowlin, A. Peterman, H. Ring, D. Seidenfeld, et al. 2017. "Cash for women's Empowerment? A Mixed-Methods Evaluation of the Government of Zambia's Child Grant Programme." *World Development* 95:55–72. https://doi.org/10.1016/j.worlddev.2017.02.017.

Boulton, J. G., P. M. Allen, and C. Bowman. 2015. *Embracing Complexity: Strategic Perspectives in an Age of Uncertainty*. Oxford: Oxford University Press.

Boxenbaum, E., and S. Jonsson. 2017. "Isomorphism, Diffusion and Decoupling: Concept Evolution and Theoretical Challenges." In *The Sage* Handbook of Organizational Institutionalism, edited by R. Greenwood, C. Oliver, T. B. Lawrence, and R. E. Meyer. London: Sage Publications.

Brousselle, A., and J. Buregeya. 2018. "Theory-Based Evaluations: Framing the Existence of a New Theory in Evaluation and the Rise of the 5th Generation." *Evaluation* 24 (2): 153–168. https://doi.org/10.1177/1356389018765487.

Camfield, L., and M. Duvendack. 2014. "Impact Evaluation – Are We 'Off the Gold standard'?" *The European Journal of Development Research* 26 (1): 1–11. https://doi.org/10.1057/ejdr.2013.42.

Cartwright, N. 2020. "Using Middle-Level Theory to Improve Programme and Evaluation Design." *CHESS Working Paper NO.2020-03*, Durham University.

Chambers, R. 2009. "So That the Poor Count More: Using Participatory Methods for Impact Evaluation." *Journal of Development Effectiveness* 1 (3): 243–246. https://doi.org/10.1080/19439340903137199.

Chambers, R. 2015. "Inclusive Rigour for Complexity." *Journal of Development Effectiveness* 7 (3): 327–335. https://doi.org/10.1080/19439342.2015.1068356.

Concern Worldwide. 2021. November. *Enabling Sustainable Graduation Out of Poverty for the Extreme Poor: An Overview of the Concern Worldwide Graduation Programme in Malawi*, Dublin. Accessed May 17, 2022. https://www.concern.net/knowledge-hub/graduation-model-and-gender-empowerment-research-project-malawi.

Copestake, J. 2014. "Credible Impact Evaluation in Complex Contexts: Confirmatory and Exploratory Approaches." *Evaluation* 20 (4): 412–427. https://doi.org/10.1177/1356389014550559.

Copestake, J. 2021. "Case and Evidence Selection for Robust Generalisation in Impact Evaluation." *Development in Practice* 31 (2): 150–160. https://doi.org/10.1080/09614524.2020.1828829.

Copestake, J., G. Davies, and F. Remnant. 2019. "Generating Credible Evidence of Social Impact Using the Qualitative Impact Protocol (QuIP): The Challenge of Positionality in Data Coding and Analysis." In *Myths, Methods and Messiness: Insights for Qualitative Research Analysis*, edited by B. Clift, G. Gore, S. Bekker, I. Batlle, K. Chudzikowskil, and J. Hatchard, 17–29. Bath: University of Bath.

Copestake, J., M. Morsink, and F. Remnant. 2019. *Attributing Development Impact: The Qualitative Impact Protocol (QuIP) Case Book*. Rugby: Practical Action.

Copestake, J., A.-M. O'Riordan, and M. Telford. 2016. "Justifying Development Financing of Small NGOs: Impact Evidence, Political Expedience and the Case of the UK Civil Society Challenge Fund." *Journal of Development Effectiveness* 8 (2): 157–170. https://doi.org/10.1080/19439342.2016.1150317.

Copestake, J., F. Remnant, C. Allan, W. van Bekkum, M. Belay, T. Goshu, P. Mvula, E. Thomas, and Z. Zerahun. 2018. "Managing Relationships in Qualitative Impact Evaluation of International Development Practice: QuIP Choreography As a Case Study." *Evaluation* 24 (2): 169–184. https://doi.org/10.1177/1356389018763243.

Creswell, J. W., and V. L. Plano Clark. 2018. *Designing and Conducting Mixed Methods Research*. 3rd ed. London: Sage Publications.

Dahler-Larsen, P. 2011. *The Evaluation Society*. Standford: Stanford University Press.

de Allegri, M., S. Brenner, C. Kambala, J. Mazalale, A. S. Muula, J. Chinkhumba, D. Wilhelm, and J. Lohmann. 2020. "Exploiting the Emergent Nature of Mixed Methods Designs: Insights from a Mixed Methods Impact Evaluation in Malawi." *Health Policy and Planning* 35 (1): 102–106. https://doi.org/10.1093/heapol/czz126.

Deaton, A., and N. Cartwright. 2018. "Understanding and Misunderstanding Randomized Controlled Trials." *Social Science and Medicine* 210:2–21. https://doi.org/10.1016/j.socscimed.2017.12.005.

Deephouse, D. L., J. Bundy, L. P. Tost, and M. C. Suchman. 2017. "Organizational Legitimacy: Six Key Questions." In *The Sage Handbook of Organizational Institutionalism*, edited by R. Greenwood, C. Oliver, T. B. Lawrence, and R. E. Meyer. London: Sage Publications.

de Milliano, M., C. Barrington, G. Angeles, and C. Gbedemah. 2021. "Crowding-Out or Crowding-In? Effects of LEAP 1000 Unconditional Cash Transfer Program on Household and Community Support Among Women in Rural Ghana." *World Development* 143:105466. https://doi.org/10.1016/j.worlddev.2021.105466.

Denzau, A. T., and D. North. 1994. "Shared Mental Models: Ideologies and Institutions." *Kyklos* 47 (1): 1–29. https://doi.org/10.1111/j.1467-6435.1994.tb02246.x.

Dunning, T. 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge: Cambridge University Press.

Eyben, R. 2013. "Uncovering the Politics of 'Evidence' and 'Results'. A Framing Paper for Development Practitioners." Accessed August 2, 2022. www.bigpushforward.net.

Fetters, M. D., and J. Molina-Azorin. 2017. "The Journal of Mixed Methods Research Starts a New Decade: The Mixed Methods Research Integration Trilogy and Its Dimensions." *Journal of Mixed Methods Research* 11 (3): 291–307. https://doi.org/10.1177/1558689817714066.

Flyvbjerg, B. 2006. "Making Organization Research Matter: Power, Values and Phronesis." In *The Sage Handbook of Organization Studies*, edited by S. R. Clegg, C. Hardy, T. B. Lawrence, and W. R. Nord, 370–387. 2nd ed. Thousand Oaks, CA, Sage: Sage Publications.

Garcia, R., and A. Zazueta. 2015. "Going Beyond Mixed Methods to Mixed Approaches: A Systems Perspective for Asking the Right Questions." *IDS Bulletin* 46 (1): 1–14. https://doi.org/10.1111/1759-5436.12119.

Gawande, A. 2008. *Better: A surgeon's Notes on Performance*. London: Profile.

Gibbs, A., J. Corboz, E. Chirwa, C. Mann, F. Karim, M. Shafiq, and A. Mecagni. 2020. "The Impacts of Combined Social and Economic Empowerment Training on Intimate Partner Violence, Depression, Gender Norms and Livelihoods Among Women: An Individually Randomised Controlled Trial and Qualitative Study in Afghanistan." *BMJ Global Health* 5 (3): e001946. https://doi.org/10.1136/bmjgh-2019-001946.

Glennerster, R., and K. Takavarasha. 2013. *Running Randomized Evaluations: A Practical Guide*. Princeton: Princeton University Press.

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

Goldstein, M., and R. Pierotti. 2020. *Mixing Qualitative and Quantitative Methods. World Bank*, DevEval Blog. Accessed August 17, 2023. https://blogs.worldbank.org/impactevaluations/mixing-qualitative-and-quantitative-methods-conversation.

Guest, G., and P. J. Fleming. 2015. "Mixed Methods Research." In *Public Health Research Methods*, edited by G. Guest and E. Namey, 581–611. Thousand Oaks CA: Sage Publications.

Haig, B., and C. Evers. 2016. *Realist Inquiry in Social Science*. Los Angeles: Sage.

Harari, Y. N. 2011. *Sapiens: A Brief History of Humankind*. New York, NY: Harper.

Hayman, R., S. King, T. Kontinen, and L. Narayanswarmy, editors. 2016. *Negotiating Knowledge: Evidence and Experience in Development NGOs*. London: Practical Action Publishing.

Hedley, E., and G. Freer. 2022. "Using Theory-Based Evaluation to Evaluate Systemic Change in a Market Systems Programme in Nepal." *IDS bulletin* 53 (1): 43–62. https://doi.org/10.19088/1968-2022.104.

Heinemann, E., A. Van Hemelrijck, and I. Guijt. 2017. July. "Getting the Most Out of Impact Evaluation for Learning, Reporting and Influence." *Insights from Piloting a Participatory Impact Assessment and Learning Approach (PIALA). IFAD Research Series*, ISBN 978-92-9072-767-5.

Hernandez, K., B. Ramalingam, and L. Wild. 2019. "Towards Evidence-Informed Adaptive Management: A Roadmap for Development and Humanitarian Organisations." *ODI Working Paper 565*.

Hewstone, M. 1989. *Causal Attribution: From Cognitive Processes to Collective Beliefs*. London: Wiley-Blackwell.

Honig, D., and L. Pritchett. 2019. "The Limits of Accountability in Education (And Far Beyond): Why More Accounting Will Rarely Solve Accountability Problems." *Working Paper 510*. Washington, DC: Center for Global Development.

Howard, N. 2022. "Towards Ethical Good Practice in Cash Transfer Trials and Their Evaluation." *Open Research Europe*. Accessed August 23, 2023. https://open-research-europe.ec.europa.eu/articles/2-12 .

Humphreys, M., and A. Jacobs. 2015. "Mixing Methods: A Bayesian Approach." *American Political Science Review* 109 (4): 653–673. https://doi.org/10.1017/S0003055415000453.

Jimenez, E., H. Waddington, N. Goel, A. Prost, A. Pullin, H. White, S. Lahiri, and A. Narain. 2018. "Mixing and Matching: Using Qualitative Methods to Improve Quantitative Impact Evaluations (IEs) and Systematic Reviews (SRs) of Development Outcomes." *Journal of Development Effectiveness* 10 (4): 400–421. https://doi.org/10.1080/19439342.2018.1534875.

Kabeer, N. 2019. "Randomized Control Trials and Qualitative Evaluations of a Multifaceted Programme for Women in Extreme Poverty: Empirical Findings and Methodological Reflections." *Journal of Human Development & Capabilities* 20 (2): 197–217. https://doi.org/10.1080/19452829.2018.1536696.

Kinstler, L. 2024. "How Poor Kenyans Became economists' Guinea Pigs. Randomised Controlled Trials Have Many Problems. They May Still Be the Best Too for Solving Poverty." *1843 Magazine, the Economist*, March, 1.

Margolies, A., E. Colantuoni, R. Morgan, A. Gelli, and L. Caufield. 2023. "The Burdens of Participation: A Mixed-Methods Study of the Effects of a Nutrition-Sensitive Agriculture Program on women's Time Use in Malawi." *World Development* 162:106122. https://doi.org/10.1016/j.worlddev.2022.106122.

Martens, B., U. Mummert, P. Murrell, and P. Seabright. 2002. *The Institutional Economics of Foreign Aid*. Cambridge: Cambridge University Press.

Maxwell, J. 2004. "Using Qualitative Methods for Causal Explanation." *Field Methods* 16 (3): 243–264. https://doi.org/10.1177/1525822X04266831.

Molecke, G., and J. Pinkse. 2017. "Accountability for Social Impact: A Bricolage Perspective on Impact Measurement in Social Enterprises." *Journal of Business Venturing* 32 (5): 550–568. https://doi.org/10.1016/j.jbusvent.2017.05.003.

Morgan, D. 2007. "Paradigms Lost and Pragmatism Regained. Methodological Implications of Combining Qualitative and Quantitative Methods." *Journal of Mixed Methods Research* 1 (1): 48–76. https://doi.org/10.1177/2345678906292462.

Morgan, J. 2019. "A Realist Alternative to Randomised Control Trials: A Bridge Not a Barrier?" *The European Journal of Development Research* 31 (2): 180–188. https://doi.org/10.1057/s41287-019-00200-y.

Moris, J. R., and J. Copestake. 1993. *Qualitative Enquiry for Rural Development : A Review*. London: Intermediate Technology Publications on Behalf of the Overseas Development Institute.

Nicholls A., J. Nicholls, and R. Paton. 2015. "Measuring social impact." *Social Finance*, edited byA. Nicholls, R. Paton and J. Emerson, 253–281. Oxford: Oxford University Press.

Olsen, W. 2019. "Bridging to Action Requires Mixed Methods, Not Only Randomised Control Trials." *The European Journal of Development Research* 31 (2): 139–162. https://doi.org/10.1057/s41287-019-00201-x.

Olsen, W., and J. Morgan. 2005. "A Critical Epistemology of Analytical Statistics: Addressing the Sceptical Realist." *Journal of the Theory of Social Behaviour* 35 (3): 255–284. https://doi.org/10.1111/j.1468-5914.2005.00279.x.

Pawson, R. 2013. *The Science of Evaluation: A Realist Manifesto*. London: Sage.

Picciotto, R., and R. Weaving. 1994. December. "A New Project Cycle for the World Bank." *Finance and Development* 31 (4): 42. https://www.imf.org/en/Publications/fandd.

Powell, S., J. Copestake, and F. Remnant. 2024. "Causal Mapping for Evaluators." *Evaluation* 30 (1): 100–119. https://doi.org/10.1177/13563890231196601.

Pritchett, L., S. Samji, and J. Mammer. 2013. "It's All About MeE: Learning in Development Projects Through Monitoring, Experiential Learning and Impact Evaluation." *Centre for Global Development, Working Paper 233*. Washington, DC.

Ramalingam, B. 2013. *Aid on the Edge of Chaos: Rethinking International Cooperation in a Complex World*. Oxford: Oxford University Press.

Ranganathan, M., M. Pichon, M. Hidrobo, H. Tambet, W. Sintayehu, S. Tadesse, and A. Buller. 2022. "Government of Ethiopia's Public Works and Complementary Programmes a Mixed Methods Study on Pathways to Reduce Intimate Partner Violence." *Social Science and Medicine* 294:114708. https://doi.org/10.1016/j.socscimed.2022.114708.

Rao, V. 2022. "Can Economics Become More Reflexive? Exploring the Potential of Mixed-Methods." *World Bank Policy Research Working Paper, No. WPS 9918*.

Rao, V., and M. Walton. 2004. *Culture and Public Action*. Washington DC: World Bank and Stanford University Press.

Ravallion, M. 2018. "Should the Randomistas (Continue To) Rule?" *Working Paper 492*. Washington, DC: Center for Global Development.

Reichardt, C. S. 2022. "The Counterfactual Definition of a Program Effect." *The American Journal of Evaluation* 43 (2): 158–174. https://doi.org/10.1177/1098214020975485.

Repko, A. F., and R. Szostak. 2021. *Interdisciplinary Research: Process and Theory*. 4th ed. Thousand Oaks CA: Sage Publications.

Rodrik, D. 2008. "The New Development Economics: We Shall Experiment, but How Shall We Learn?" *Research Working Paper 08-055*. Harvard Kennedy School: John F. Kennedy School of Government.

Rogers, P. 2020. December. "Real-Time Evaluation. Monitoring and Evaluation for Adaptive Management." *Working Paper Series, No.4*, Accessed September 19, 2023. www.betterevaluation.

Schwandt, T., and E. Gates. 2021. *Evaluating and Valuing in Social Research*. New York: The Guildford Press.

Senge, P. 1990. *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Doubleday.

Stern, E., N. Stame, N. Mayne, K. Forss, R. Davies, and B. Befani. 2012. "Broadening the Range of Designs and Methods for Impact Evaluations." *Working Paper No. 3*. London: Department for International Development.

Tashakkori, A., and J. Creswell. 2007. "Editorial: The New Era of Mixed Methods." *Journal of Mixed Methods Research* 1 (1): 4–7. https://doi.org/10.1177/2345678906293042.

Warren, E., G. Melendez-Torres, and C. Bonell. 2022. "Are Realist Randomised Controlled Trials Possible? A Reflection on the INCLUSIVE Evaluation of a Whole-School Bullying Prevention Application." *Trials* 23 (1): 82. https://doi.org/10.1186/s13063-021-05976-1.

Webster, J., J. Exley, J. Copestake, R. Davies, and J. Hargreaves. 2018. "Timely Evaluation in International Development." *Journal of Development Effectiveness* 10 (4): 482–508. https://doi.org/10.1080/19439342.2018.1543345.

White, H. 2011. "Achieving High-Quality Impact Evaluation Design Through Mixed Methods: The Case of Infrastructure." *Journal of Development Effectiveness* 3 (1): 131–144. https://doi.org/10.1080/19439342.2010.547588.

White, H. 2019. "Comment: The Twenty-First Century Experimenting Society: The Four Waves of the Evidence Revolution." *Palgrave Communications* 5 (1): 47. https://doi.org/10.1057/s41599-019-0253-6.

White, S. 2015. "Qualitative Perspectives on the Impact Evaluation of girls' Empowerment in Bangladesh." *Journal of Development Effectiveness* 7 (2): 127–145.

White, H., and D. Phillips. 2012. "Addressing Attribution of Cause and Effect in Small N Impact Evaluations: Towards an Integrated Framework." *International Initiative for Impact Evaluation, Working Paper 15*, Accessed February 26, 2024. www.3ieimpact.org.

White, H., and D. A. Raitzer. 2017. *Impact Evaluation of Development Interventions: A Practical Guide*. Manila: Asian Development Bank.

Woolcock, M. 2009. "Towards a Plurality of Methods in Project Evaluation: A Contextualised Approach to Understanding Impact Trajectories and Efficacy." *Journal of Development Effectiveness* 1 (1): 1–14. https://doi.org/10.1080/19439340902727719.

Woolcock, M. 2019. "Reasons for Using Mixed Methods in the Evaluation of Complex Projects." *Faculty Working Paper, No.348*. Center for International Development at Harvard University.

Appendix. List of QuIP studies conducted by Bath SDR (2017 – 2023)

Agriculture and rural livelihood promotion

Commissioner (and main donor)	Country	Year fini-shed	QuIP inter-views	QuIP focus groups	Brief description of the intervention being evaluated
Acumen	India	2016	64	0	Impact investment in dairy farming
Diageo Ltd	Ethiopia	2016	48	8	Malt barley cultivation*#
Oxfam UK	Ethiopia	2016	58	8	Value chain promotion: cooperative coffee processing
Self Help Africa	Kenya	2016	24	4	Value chain promotion: production and bulking of cassava, pigeon peas and green gram
Diageo Ltd	Uganda	2017	48	8	Provision of seeds, farming equipment and training to improve sorghum, cassava & malt barley production
Self Help Africa	Burkina Faso	2017	24	4	Promoting sorghum and millet cultivation and marketing
Self Help Africa	Zambia	2017	46	4	Integrated rural livelihoods programme: agricultural productivity, nutrition and health education
Tree Aid	Ghana	2017	24	4	Strengthening the shea bean value chain for women
ITAD Ltd (DFID)	Nepal	2018	96	12	Making markets work for the poor (support for producers of vegetables, dairy, pigs & ginger)
Trinity College Dublin – TCD (Concern)	Malawi	2018	24	2	Cash 'plus' graduation out of poverty (pilot)
Aga Khan Fdn (USAID)	Pakistan	2019	24	4	Irrigation infrastructure support and agribusiness training
Aga Khan Support Fdn	Pakistan	2019	48	8	Agriculture and nutrition support for most food insecure households*
AgDevCo Ltd (Mastercard Fdn)	Uganda	2019	48	8	Smallholder farmer access to microloans via Village Savings and Loans Associations*
Diageo Ltd	Came-roon	2019	48	8	Training and technical assistance in WASH, farming and nutrition – focus on empowerment of women
Send a Cow (DFID)	Ethiopia	2019	36	6	Training farmer self-help groups in sustainable agriculture and gender awareness
AgDevCo Ltd (FCDO)	Uganda	2021	36	4	Capacity building and job opportunities for women in agribusiness
Diageo Ltd	Kenya	2021	48	8	Agricultural training and input supply under contract farming
Opportunity International (DFID)	Ghana	2021	48	8	Agricultural capacity building and access to financial services for women (endline)
TCD (Concern)	Malawi	2021	48	8	Cash 'plus' graduation out of poverty (midline)
PDA Associates (AgDevCo Ltd)	Ghana	2022	24	4	Impact investment into the Babator Irrigated Farming Hub (midline impact evaluation)
Fairtrade Foundation	Ivory Coast	2022	42	0	Membership of fairtrade cocoa cooperatives supplying Ben & Jerry's
Self Help Africa (EU)	Kenya	2022	24	4	Vegetable farming: pre- and post-harvest storage, packing and management
Self Help Africa (EU)	Kenya	2022	24	4	Training in climate smart agricultural practices for pyrethrum farming

(Continued).

Commissioner (and main donor)	Country	Year finished	QuIP inter-views	QuIP focus groups	Brief description of the intervention being evaluated
TCD (Concern)	Malawi	2022	48	8	Cash 'plus' graduation out of poverty (endline)
Self Help Africa (EU)	Kenya	2023	48	8	Access to agricultural inputs, training and reliable purchasing contracts for cotton and bean farmers

Health promotion (including training medical and health workers)

Commissioner (and main donor)	Country	Year finished	QuIP inter-views	QuIP focus groups	Brief description of the intervention being evaluated
Save the Children	Ethiopia	2017	24	4	Early famine response
Save the Children (Irish Aid)	Tanzania	2017	48	8	Integrated agriculture, nutrition, WASH and childcare*#
SEED Global Health (Peace Corps)	Uganda	2017	24	8	Contribution of expert volunteers to medical training#
SEED Global Health (Peace Corps)	Malawi	2017	36	7	Contribution of expert volunteers to medical training#
SEED Global Health (Peace Corps)	Tanzania	2017	24	4	Contribution of expert volunteers to medical training#
The C&A Foundation	Mexico	2017	33	4	Promoting health and better working conditions among garment factory workers#
Aga Khan University	Uganda	2018	14	0	Career development of graduate midwives
Mannion Daniels (DFID)	Multi-country	2018	24	0	Capacity building for NGOs promoting sexual and reproductive health rights
Rutgers International	Kenya	2018	24	4	Sexual and reproductive health rights awareness and support services
Aga Khan Foundation	Tajik-istan	2019	36	6	Training and technical support on food security and nutrition
Rutgers International	Uganda	2019	24	4	Training young people to become community health entrepreneurs
Save the Children (DFID, Helen Keller)	Moza-mbique	2019	48	8	Integrated agriculture, nutrition, WASH, and childcare
Rutgers	Indo-nesia	2020	48	12	Teacher-led intervention for 12–15 year-old children to support healthy and positive sexual development
Girl Effect (Nike Fndtn)	Rwanda	2021	48	8	Media platform to support and inform girls about their sexual and reproductive health behaviour
Juntos! (La Caixa & Aga Khan Fdn)	Moza-mbique	2021	22	0	Capacity building of advocacy for civil society organisations
Save the Children	Zimb-abwe	2021	24	4	Integrated agriculture, nutrition, WASH and childcare*
UNICEF	Serbia	2023	36	0	Changes to a health mediator service to improve Roma community access to formal medical services for infants

other

Commissioner (and main donor)	Country	Year finished	QuIP inter-views	QuIP focus groups	Brief description of the intervention being evaluated
Acumen	India	2017	64	0	Impact investment in training for women to set up businesses (e.g. beauty parlours, tailoring)
Habitat for Humanity	India	2017	72	4	Microfinance for housing improvements*#
Tearfund	Uganda	2017	48	8	Faith based community mobilisation*#
Voscur UK	UK	2017	24	0	Council support for community-based organisations in Bristol
Tearfund	Sierra Leone	2018	48	8	Faith based community mobilisation*
Tearfund	Bolivia	2018	48	8	Faith based community mobilisation*
OPM (Master Card Fdn)	Ghana	2019	24	4	Investment in technology to link formal financial services with informal savings mechanisms
Voscur (Bristol City Council)	UK	2019	24	0	Council support for community organisations in Bristol*
OPM (Master Card Fdn)	Tanzania	2020	24	4	Investment in technology to link formal financial services with informal savings mechanisms*
OPM (Master Card Fdn)	Zambia	2020	24	4	Investment in technology to link formal financial services with informal savings mechanisms*
Opportunity International (DFID)	Ghana	2020	48	8	Agricultural capacity building and access to financial services for women (midline)
Power to Change (Nat. Lottery)	UK	2020	24	0	Training and support for community enterprise development*
TripleLine Ltd (DFID)	Ghana	2020	20	0	Governance of timber extraction
University of Bath	Sierra Leone	2020	27	0	Regulation of artisanal mining
EduFinance	Kenya	2021	48	0	Microfinance for low-cost private schooling*
OPM (Master Card Fdn)	Tanzania	2021	96	8	Investment in technology to link formal financial services with informal savings mechanisms*
Power to Change (Big Lottery Fund)	UK	2021	24	0	Training and support for community enterprise development
Kantar Public (Home Office)	UK	2022	48	6	Police and community interventions to improve safety for women
Kantar Public (Home Office)	UK	2022	23	0	Police and community interventions to improve safety for women in the night-time economy
OPM (Master Card Fdn)	Ghana	2022	24	0	Investment in technology to link formal financial services with informal savings mechanisms*
Opportunity International (Global Affairs Canada)	Ghana	2022	50	0	Improved financial services for women owners of small and medium enterprises
Kantar Public (DCMS)	UK	2023	40	0	Distribution of hypothecated value-added tax on sanitary products to NGOs supporting women and girls*

Sources.

The table lists all QuIP studies in which Bath SDR Ltd was contracted by the named commissioner and was compiled from its archives. It excludes QuIP studies carried out independently by other individuals or organisations.

*Original or summary report available on the Bath SDR website - https://bathsdr.org/resources/?_resource_types=example-reports.

#Written up as a case study chapter in Copestake et al. (2019a).