Mixed methods impact evaluation in international development: distinguishing between ‘quant-led’ and ‘qual-led’ approaches.

James Copestake.

University of Bath, UK.

Pre-submission draft 6 October 2023

Abstract

Despite being widely endorsed for more than two decades, the practice of mixed methods impact evaluation (MMIE) remains confused. This paper suggests greater clarity can be achieved by distinguishing between ‘quant-led’ and ‘qual-led’ approaches to MMIE, both of which incorporate quantitative and qualitative steps. After describing each approach, it draws on published studies and direct experience of using the Qualitative Impact Protocol (QuIP) within both approaches to compare them. While technically appropriate to different kinds of intervention, it suggests that path-dependent preference constraints and the interests of evaluation commissioners and researchers also influences the choice between them, as well explaining differences in how widely findings are disseminated. The paper is mainly intended to be of practical relevance to those planning, conducting, and reviewing both approaches to MMIE, but also relates to wider concerns about power and the ethics of knowledge production and distribution.

Key words

Adaptive management, impact evaluation; mixed methods research, qualitative and quantitative research methods

Introduction

Mixed method 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 et al. 2010). The general case for combining quantitative and qualitative approaches mainly rests on two arguments - that the strengths of each approach can mitigate the weaknesses of the other, and that their convergent integration can add to the overall credibility of findings (e.g. Woolcock 2019, 4). But 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 (White 2011, Bamberger 2015, White 2015, Jimenez *et al.* 2018, Kabeer 2019). This paper proceeds from the premise that one route to realising the potential value of MMIE is to increase understanding of the different ways in which it is currently conducted.

The paper focuses on impact evaluation, rather than other forms of research in its central concern for identifying outcomes of a specific intervention, whether a time-bound project or experiment, or a more open-ended programme or policy. It builds on the distinction between variance-based and process theory-based approaches to causal attribution to identify two dominant approaches to MMIE in international development. The first is labelled 'quant-led' and relies mainly on variance-based impact evaluation, but its use also entails performing qualitative tasks. It also increasingly accommodates process theory-based attribution in a complementary, if generally subordinate way. The second is labelled 'qual-led' and relies mainly on process theory-based attribution, but also incorporates collection and use of quantitative data. For example, realist evaluations often use quantitative data to identify variation in outcomes and contexts of interventions, but rely on process theory-based attribution to identify causal mechanisms (Pawson, 2013).

Having clarified the difference between two dominant approaches to MMIE, I draw on published literature and direct experience with using the Qualitative Impact Protocol (QuIP) within both approaches to discuss their methodological strengths and weaknesses.[1] This addresses the main purpose of the paper - to be of practical use to those planning, conducting, and reviewing MMIEs by aiding understanding of the range of approaches available. Reflection on the methodological strengths and weaknesses of the two approaches also leads into a discussion of the preference constraints and interests of commissioners and researchers, as well as the wider political economy of knowledge production. Quant-led approaches are congruent with a world view that favours relatively replicable, technical, and linear models of development intervention, whereas qual-led approaches fit better with a view of development that is more path-dependent, social, and

[1] The QuIP is a qual-led approach to MMIE based on collecting, coding, analysing and mapping narrative accounts of the causal drivers of change (Copestake *et al.* 2018, 2019a, 2019b, Copestake 2021).

complex. The two approaches generate different kinds of evidence, but also serve distinct interests, and it is not obvious that one is necessarily more useful in bringing evidence to bear on complex development issues than the other. This remains uncertain and warrants further research, as do questions about whose interests are best served by current norms for withholding and disseminating findings from the two approaches.

Conceptual framing

A brief excursion into concepts and definitions can be justified by confusion arising from ambiguity in the use of the core concepts of 'quantitative' and 'qualitative' that underpin the idea of mixed methods.[2] I start by distinguishing between two approaches to causal attribution - the challenge that lies at the heart of impact evaluation - then develop a broader framework for thinking in a more granular way about quantitative and qualitative aspects of different approaches to MMIE. The discussion thereby negotiates three quite different ways of thinking about the qual-quant dichotomy - as two distinct cultures, as two sets of research methods, or 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**), with the bold type indicating the **X**, **Y** and **Z** are vectors of factors. This nomenclature can be used to draw the distinction (taken from Maxwell 2004) between successionist or "variance" based (primarily quantitative) and generative or "process theory" based (primarily qualitative) approaches to causal attribution. For researchers working with the variance-based approach, this entails identifying a counterfactual - or what *would* have happened to **Y** if **X** had not happened, with **Z** remaining the same (Dunning, 2012; Glennerster and Takavarasha, 2013; 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

[2] 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). This is a fuzzy distinction (as explained later in the paper) but excludes discussion of many methods used mostly in isolation. Tashakkori and Creswell (2007) and Creswell and Plano Clark (2018) elaborate further on the definition of mixed method research in general.

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**. Randomized controlled trials (RCTs), natural experiments and other quasi-experimental methods offer a range of solutions to this problem.

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 the process of how change happens. The challenge facing researchers is to render this understanding explicit in a way that can be subjected to critical scrutiny and contribute to useful generalisation.[3] An advantage of this approach 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. But the evidence collected is conceptually fuzzier, and generalisation is messier than with a more quantitative approach because data is not collected to fit a predetermined conceptual framework or coding pattern (Powell *et al.*, forthcoming). 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 (Stern *et al.* 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 are useful in explaining the world by filling the vast chasm between universal laws, and theories of change for interventions unique to one time and place (Cartwright 2020).

A 'belt and braces' approach to MMIE would make 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 the normative model that underpins how many researchers think about MMIE, with the first (variance) approach classified as a quantitative method, and the second (process theory) approach as a qualitative method. This conflation of quant and qual with attribution methods is also

[3] 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. Raichardt (2022) offers a comprehensive review of different forms of counterfactual thinking.

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).[4] Distinguishing between relatively self-contained quantitative and qualitative methods also reflects a wider tendency to do so across the social sciences. 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 observes how the paradigmatic shift to methodological pluralism has eroded these distinctions: integrating induction and deduction through processes of abduction and retroduction; 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 its efficient collection, storage, and manipulation in numerical form. More qualitative approaches, in contrast, are based on *delayed codification*. This delay makes it harder to handle large amounts of data numerically but requires fewer assumptions about what data to 'admit' and in what form, thereby increasing the range of possible findings (Moris and Copestake 1993). This distinction can be applied to tasks *within* methods as well as being a criterion for distinguishing *between*

[4] 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 democratize and enrich otherwise ethically dubious processes of data extraction.

methods, thereby permitting more granular distinctions between mixed method research designs.

With impact evaluation, an additional source of complexity arises from the need to synchronise these tasks within the flow of activity being evaluated (Webster *et al.* 2018). A simple way to illustrate this is distinguish between tasks carried out before (t=1), during (t=2) and after (t=3) phases of an intervention.[5] Adapting the nomenclature favoured by Creswell and Plano Clark (2018), interactions between quantitative and qualitative tasks can be identified both *within* and *between* intervention phases, as illustrated in Table 1. This highlights the existence of many possible designs of mixed method impact evaluation, with additional variation arising from the relative *weight* and *purpose* of each component and its interaction with other tasks (Guest and Fleming, 2015).

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

Quant-qual interactions *within* phases	Quant-qual interactions *between* phases
quant1 -> qual1	quant1 -> qual2
qual1 -> quant1	qual1 -> quant2
quant2 -> qual2	quant1 -> qual3
qual2 -> quant2	qual1 -> quant3
quant3 -> qual3	quant2 -> qual3
qual3 -> quant3	qual2 -> quant 3

A further elaboration of this framework would be to distinguish between evaluative tasks required for (a) framing and design an impact evaluation, (b) data collection, and (c) data analysis and use. There is a tendency for the first of these to take place mostly *before* the intervention (t=1), the second *during* its implementation (t=2), and the third *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 require pre-intervention data collection in the form of a baseline survey, so as not to have to rely on respondent recall; and natural experiments depend less on having to

[5] Of course, evaluations must often also synchronise with multi-stage interventions, including piloting and mainstreaming (Picciotto and Weaving 1994), unplanned interruptions and adjustments.

synchronise the intervention with its evaluation (Dunning, 2012), whereas RCTs require prospective co-design of the intervention and its evaluation to permit random assignment of treatments.

Analysis

The previous section highlighted the existence of many possible designs for MMIE. This section uses this framework and draws upon published literature on MMIE and direct experience with use of the QuIP to distinguish between just two dominant approaches, referred to as 'quant-led' and 'qual-led', where the core difference arises from the weight ascribed to primarily quantitative (variance based) and qualitative (process theory) based causal attribution.

Quant-led approaches to MMIE

At its simplest this can be broken down into three stages: 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. This model primarily generalises inductively from a systematic review of forty impact evaluation studies conducted by Jimenez *et al.* (2018).[6] It was also tested against more recently published examples (de Allegri *et al.* 2020, de Milliano *et al.* 2021, Margolies *et al*. 2021, and Ranganathan *et al.* 2022), and incorporates insights from the more critical perspectives of quant-led MMIE by White (2015) and Kabeer (2019). Table 2 elaborates on key tasks and their interactions across the three intervention phases.

[6] 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.

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.

In the first phase, the main qualitative task is to inform design of the variance-based impact evaluation. This includes contextual analysis, refining the theory of change informing evaluation design, defining key concepts, identifying measurable indicators for them, and pilot testing research instruments (White 2011; Garcia and Zazueta 2015). Once key research questions are agreed, then statistical power calculations can 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, but 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 – precise estimates of the magnitude of impact on preselected indicators for one, and uncertain levels of insight into the causal drivers of these effects and their perceived importance to different stakeholders, in the other.

The baseline survey provides the foundation for quantitative impact assessment if followed up 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 controlling also for variation in **Z**.[7] 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.[8] 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** established quantitatively. 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.[9]

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 minimal studies relegate qualitative tasks to mapping the context and assessing implementation fidelity, rather than contributing to causal inference. More comprehensive studies

[7] Many studies also include a 'Quant2' or mid-line survey through which intermediate impact can be assessed quantitatively before the intervention ends. This can be regarded a hybrid model between the two considered here, to the extent that it support adaptive mid-course adjustments.
[8] 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.
[9] For an example, see Bonilla *et al.* (2017). This draws on a qualitative strand to suggest and illustrate causal mechanisms consistent with quantitative findings, to identify avenues for quantitative analysis of heterogeneous impact, and to question the robustness of key outcome indicators of women's empowerment.

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.

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 core cost-benefit questions. White (2015) also highlights insufficient involvement of experienced qualitative researchers in the design and management of such studies, pointing to a need for discussion that goes beyond what different approaches can and cannot deliver in theory (see below).

Qual-led approaches to MMIE

A contrasting approach to the above centres on qualitative enquiry informed by concurrent quantitative monitoring. At its simplest, it looks like this:

(quant2←→QUAL2)

The approach is particularly suited to evaluation of open-ended programmes and policies, but it can easily be extended for use with time-bound projects too. Initial qualitative activities include widely 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 rather than impact evaluation, but nevertheless provide an essential quantitative foundation for it. Qualitative impact evaluation builds on it by providing additional feedback to enable internal and external stakeholders to assess the causal processes behind observed trends and changes. The decision about how much additional evidence is needed, when and why, is partially institutionalised but also adjusts reflexively in response to specific questions and crises as they arise. This resonates with both an opportunistic, *bricolage* approach to MMIE (Aston and Apgar 2022) and a more formal Bayesian approach (Humphreys and Jacobs 2015). It is also consistent with pragmatically augmenting the causal judgements integral to performance management by commissioning more formal impact evaluation studies, using both process theory-led impact evaluation methods such as contribution analysis, outcome harvesting, process tracing, realist evaluation (Copestake *et al.* 2019b, ch.2), and participatory approaches (Chambers, 2009; Heinemann *et al.* 2017). Indeed variance-based impact evaluation also challenge and refine an organisation's understanding of its impact in this way.

While consistent with a more adaptive and complexity-informed view of development practice (e.g. Hernandez *et al.* 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 still 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.[10] 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 in monitoring indicators informs specification and focus of qualitative impact evaluation – see Table 3.

Table 3. The qual-led impact evaluation approach (at its simplest)

Phase	**During (t=2)**
Main evaluative activities	Process evaluation. Timely learning. Adaptive management.
Mainly qualitative components	Data collection and analysis bricolage. Process-theory based causal attribution. 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 case selection for process-theory based impact evaluation studies. 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.

[10] Gawande (2008) provides powerful insights into this way of thinking, while Eyben (2013) and Honig and Pritchett (2019) explore traps arising from accountability based too strictly on rigid quantitative targets.

Further insight into this approach draws on experience of being commissioned to conduct 64 discrete impact evaluation studies using the Qualitative Impact Protocol (QuIP) 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. Of these, 21 were primarily concerned with supporting rural livelihoods, 14 with health promotion, ten with community mobilisation, and others with financial inclusion, market development, education, and crime prevention. They all focused on assessing how far selected programmes and projects were delivering intended benefits to defined target groups who included farming households, factory employees, students, disadvantaged clients and users of public services, and community level organisations. The typical study comprised 36 semi-structured interviews and four focus group discussions through which participants were given the opportunity to share their own perception of drivers of change in selected domains of their individual and collective wellbeing over a specified period. The data was coded and analysed using causal qualitative data analysis and causal mapping (Powell *et al.* forthcoming). It was then presented back to commissioning organisations to inform discussion of how far respondents' experience of change, and perception of its causal drivers, were consistent with commissioners' prior theories and expectations.

In a small number of cases the QuIP was closely integrated with quantitative impact evaluation in line with the quant-led approach already discussed. For example, evaluation of a pilot cash transfer programme in Malawi combined three rounds of QuIP studies over three years alongside an RCT (Concern Worldwide 2021). Other studies were more loosely triangulated sequentially or in parallel with quasi-experimental impact evaluation, including one study, in Tanzania, commissioned to fill a gap created by the failure of an RCT (Copestake, *et al.* 2019b). This created opportunities for selection of respondents for interview from the samples of project participants already identified through more extensive baseline surveys. However, the practical difficulty of achieving the ideal of fully integrated 'belt-and-braces' MMIE is illustrated by the our failure - even once - to select respondents purposively using measured changes in outcome indicators based on prior longitudinal surveys.

In most instances QuIP studies were not conducted alongside quant-led impact evaluation but were able to draw on routine quantitative monitoring of intended beneficiaries. In the case of microfinance institutions, for example, this covered clients' basic socio-economic characteristics plus information on their saving and borrowing activities. Using such data for case selection entailed addressing data protection and data management issues, complicated by weak connections between the operational staff managing such data and staff commissioning the evaluation, who were often employed by different organisations. The pay-off to overcoming these problems was enhanced confidence that findings reflected important sources of variation across the wider population of intended beneficiaries.

Differences in the positionality and interests of staff also affected integration of data analysis, interpretation, and use. While some organisations invested in capacity to conduct QuIP studies internally, most were sub-contracted to independent research teams. Their remit included delivering written findings, but rarely extended to participating in follow-up activities through which their significance could be assessed alongside other evidence of impact available within the commissioning organisation. Hence a key qual-quant interaction – integrated interpretation of QuIP findings and internal assessment of drivers of change informed by quantitative monitoring - remained hidden to external audiences.

Discussion

The previous section juxtaposed two contrasting approaches to MMIE, starting with a quant-led approach centred on variance-based attribution, supported by qualitative contextualisation and design, and supplemented (often weakly) with process theory-based attribution to help explain findings. We then reviewed a qual-led approach that combines quantitative monitoring with process theory-based attribution. We noted that the quant-led approach is more associated with more projectized and linear interventions, whereas the qual-led approach suits adaptive management of more open-ended programmes and policies. Quant-led MMIE can provide the clear, credible, and precise evidence of the magnitude of attributable impact that commissioners demand. Qual-led MMIE can inform wider reflection on the relevance of interventions, pick up unexpected causes and effects and enrich critical analysis of the theory underpinning interventions. However, the aim of this paper is not to attempt an overall judgement on the appropriateness of quant-led or qual-led approaches to different purposes and contexts. Instead, this section uses the distinction between the two approaches to develop a more speculative analysis of the political economy of MMIE. This emphasises the unavoidably socio-political aspects of the design of impact evaluation. More than a technical problem of evidence collection, it is also embedded in competitive and collaborative processes of securing funding - both for commissioners seeking to legitimise their activities, and for the researchers they choose to employ. Deployment of methodological discourse is a currency in these struggles.

Of the two, it is the quant-led approach 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 transformational changes in their knowledge, attitude, and behaviour (Banerjee and Duflo 2012). The growth of such projects also reflects the attraction to international donors of relatively technocratic interventions with measurable impact goals that can be replicated and scaled-up across diverse contexts, using projectized results-based 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 evaluation, with 'large-n' or variance-based methodologies capable of delivering precise estimates of impact on predetermined indicators that can relatively easily be linked to the SDGs.

Alongside this kind of intervention are more flexible modalities based on adaptive management to address more complex political, institutional, and structural development problems (Andrews *et al.* 2012; Ramalingam, 2013; Boulton *et al.* 2015). These have been associated with support for evaluative practices better attuned to uncertain impact trajectories (Woolcock, 2009), identification of unintended consequences (Bamberger *et al.* 2016) and to informing timely programme adjustments (Webster *et al.* 2018). Contextual and operational complexity also explains increased interest among development professionals in alternative approaches to impact evaluation (Stern *et al.* 2012; Brousselle and Buregeya 2018). The qual-led approach to 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 enlightened planners first select different forms of development intervention to address higher level goals, and then select appropriate approaches to MMIE to fit. However, evaluation is affected not only by task-specific 'best practices' but also by commissioners' wider interests and preferences (Martens *et al.* 2002). Commissioners' interest in evidence of impact – or lack of it - also depends on the importance they attach to it compared to the political "warm glow" of being seen to do good works (Copestake *et al.* 2016). Their methodological choices may also be limited by preference constraints and limited navigational capacity arising from the their own technical training in research methods (Rao and Walton 2004) and by dominant disciplinary norms. For example, a strong commitment to empiricism may in part reflect commissioners' unfamiliarity with realist and complexity-informed understanding of the social sciences (Bhaskar 2016; Boulton 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 twenty-year growth in donor investment in RCTs after 2003 (Camfield and Duvendack 2014, White 2019, Bédécarrats *et al.* 2020; Howard, 2022). This can be attributed in part to their fit with the technocratic genre of development projects described above. In addition, advocates of RCTs effectively 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. The critical pushback that RCTs attracted (Rodrik 2008, Basu 2014, Deaton and Cartwright 2018, Ravallion 2018) 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" (White 2019) and with a simplistic view of the transferability of natural science empiricism to the social sciences. Either way, having persuaded many evaluation commissioners that RCTs amounted to a "gold standard" proponents of them are well placed to endorse a supporting use of theory-based methods in a subordinate role within quant-led approaches to MMIE.

This brief review of the debate over RCTs should also be viewed in the context of a much older and deeper perspective on MMIE entrenched in development practice. Molecke and Pinkse (2017) distinguish between four practical arguments for discounting just about *any* source of evidence about impact: key outcomes can't 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. It does not follow from such doubts that development pragmatists who hold these views also reject the reality or importance of impact entirely, but it does incline them against investing in formalised impact evaluation more than they are forced to. Confronting high levels of complexity and uncertainty also puts a premium on experience-based wisdom or *phronesis* (Flybjerg 2006, Pritchett *et al.* 2013) and on relying for evidence of impact on multiple and grounded sources, particularly those based on direct personal observation and trusted relationships (Nicholls *et al*. 2015, 276). While more likely to advocate qual-led than quant-led approaches to MMIE, radical advocates of this view are suspicious of almost *any* formal approach to producing evidence of impact that claims to trump their insider understanding.

Beyond personal taste or temperament, distrust of formal 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 Shon 1978, Senge 1990, Denzau and North 1994). A central role of *ABC*'s leadership is to manage tensions arising from the tendency for theory espoused at the top of organization to become decoupled from everyday practices and theories-in-use lower down the organization and across collaborating organisations (Boxenbaum & Jonsson, 2017). In this context, formal evaluation can be viewed as a form of political deliberation that has both possible instrumental value (to facilitate learning, demonstrate goal achievement, account to stakeholders), but also potential for misuse within wider struggles over organizational reputations and legitimacy (Alkin and King 2016, 2017; Deephouse *et al.* 2017).

This brief excursion into organisational institutionalism illustrates why it is understandable that leaders of development agencies are careful about both the commissioning of impact evaluations and dissemination of their findings. Even if *ABC* invests in a balanced mix of internal and external evaluative activities, its internal processes of learning remain largely invisible to external stakeholders. Experience with the 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 quant-led MMIE are also open to reputational damage if they agree from the outset to independent publication of the findings, but this may be a risk worth taking when linked to funding of large-scale interventions.

A consequence of the closer association between qual-led MMIE and adaptive approaches to 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 projectized 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 impact 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 possibly also reflects greater agreement among quant-led providers about quality standards and benchmarks. Contributors to qual-led MMIE may also be content with gaining privileged insider influence by agreeing to contracts that strictly curtail what they can disseminate more widely. 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 based attribution within quant-led approaches and variance based attribution within qual-led approaches - e.g. through realist RCTs.[11] There is also scope for building stronger standards for discrete qualitative impact evaluation studies (such as the QuIP) to facilitate their wider publication, while they remain, and are seen 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), not least in development practice. This paper has sought to counter this in four ways. First, it has emphasised the existence of two radically different approaches to addressing the problem of causal attribution. Second, it has offered a framework for more fine-grained analysis of the use of qualitative and quantitative methods within MMIE and used it to distinguish between 'quant-led' and 'qual-led' approaches to it. Third, it has explored how design of MMIE depends on more than technical design considerations - how political economy and the path dependent preferences and interests of commissioners and researchers also matter. Fourth, it has suggested that asymmetry in norms affecting the dissemination 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.

Acknowledgements

This paper is a spin-off from ongoing action research based on use of the QuIP and I am grateful to many participants in QuIP studies across the world. Rebekah Avard, Will Airey, Tara Bedi, Gary Goertz, Hannah Mishan, Jennifer Golan, Michael King, Peter Mvula, Fiona Remnant, Steve Powell and Sarah White all provided helpful comments on earlier drafts. The paper also benefited from my participation in activities organised by CEDIL (The Centre of Excellence for Development Impact and Learning) funded by the UK Foreign, Commonwealth and Development Office.

Declaration of interest

There are no conflicts of interest to declare.

[11] For discussion of the scope for econometric analysis within realist research and evaluation see Morgan (2019), Neilsen *et al*. (2023) and Warren *et al*. (2022).

References

Alkin, M.C., King, J.A. (2016) The historical development of evaluation use. *American Journal of Evaluation*, 37(4):568-79.

Alkin, M.C., King, J.A. (2017) Definitions of evaluation use and misuse, evaluation influence and factors affecting use. *American Journal of Evaluation*, 38(3):434-450.

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

Argyris, C., Schon, D. (1978) *Organizational learning: a theory of action perspective*. Reading, MA: Addison-Wesley.

Aston, T., Apgar, M. (2022) *The art and craft of bricolage in evaluation*. Centre for Development Impact Practice Paper, No.24. www.idea.acv.uk/cdi

Basu, K. (2014) Randomisation, Causality and the Role of Reasoned Intuition, *Oxford Development Studies*, 42(4):455-472, DOI: 10.1080/13600818.2014.961414

Bamberger, M (2015) Innovations in the use of mixed methods in real-world evaluation, *Journal of Development Effectiveness*, 7(3):317-326, DOI: 10.1080/19439342.2015.1068832

Bamberger, M., Tarsilla, M., Hesse-Biber, S. (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.

Bamberger, M., Rao, V., Woolcock, M. (2010) *Using mixed methods in monitoring and evaluation: experiences from international development*. World Bank, Policy Research Working Paper, 5245.

Banerjee, A.V., Duflo, E. (2012) *Poor Economics: barefoot hedge-fund managers, DIY doctors and the surprising truth about life on less than $1 a day*. London: Penguin Books.

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

Bhaskar, R. (2016) *Enlightened common sense: the philosophy of critical realism*. London: Routledge.

Bonilla, J., Zarzur, R.C., Handa, S. 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

Boulton, J.G., Allen, P.M., Bowman, C. (2015) *Embracing complexity: strategic perspectives in an age of uncertainty*. Oxford University Press.

Boxenbaum, E., Jonsson, S. (2017) Isomorphism, diffusion and decoupling: concept evolution and theoretical challenges. In Greenwood, R., Oliver, C., Lawrence, T.B., Meyer, R.E. editors, The SAGE *Handbook of organizational institutionalism*. SAGE.

Brousselle, A., Buregeya, J. (2018) Theory-based evaluations: Framing the existence of a new theory in evaluation and the rise of the 5th generation. *Evaluation.,* 24(2), pp. 153-168.

Camfield, L., Duvendack, M. (2014) Impact evaluation – are we 'off the gold standard'? *European Journal of Development Research*, 26(1):1-11.

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.

Concern Worldwide (2021) *Enabling sustainable graduation out of poverty for the extreme poor: an overview of the Concern Worldwide Graduation Programme in Malawi*. November. Dublin.

https://www.concern.net/knowledge-hub/graduation-model-and-gender-empowerment-research-project-malawi [accessed 17 May 2022]

Copestake, J., O'Riordan, A-M, Telford, M (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.

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

Copestake, J., Morsink, M., Remnant, F. (2019a) *Attributing development impact: the qualitative impact protocol (QuIP) case book*. Rugby: Practical Action.

Copestake, J., Davies, G., Remnant, F. (2019b) 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. (2021) Case and evidence selection for robust generalisation in impact evaluation. *Development in Practice.* 31(2): 150-160.

Creswell, J.W., Plano Clark, V.L. (2018) *Designing and conducting mixed methods research*. Sage Publications. 3rd Edition.

Dahler-Larsen, P (2011) *The evaluation society*. Standford: Stanford University Press.

de Allegri, M., Brenner, S., Kambala, C., Mazalale, J., Muula, A.S., Chinkhumba, J., Wilhelm, D. and Lohmann, J. (2020) Exploiting the emergent nature of mixed methods designs: Insights from a mixed methods impact evaluation in Malawi. *Health Policy and Planning,* 35(1), pp. 102-106.

de Milliano, M., Barrington, C., Angeles, G. and Gbedemah, C., 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.

Deaton, A., Cartwright, N. (2018) Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine*, 210:2-21.

Deephouse, D.L., Bundy, J., Tost, L.P., Suchman, M.C. (2017) Organizational legitimacy: six key questions. In Greenwood, R., Oliver, C., Lawrence, T.B., Meyer, R.E. editors, *The SAGE Handbook of organizational institutionalism*. SAGE.

Denzau, A.T., North, D. (1994) Shared mental models: ideologies and institutions. *Kyklos*, 47(1):1-29.

Dunning, T. (2012) *Natural experiments in the social sciences: a design-based approach*. Cambridge University Press.

Eyben, R. (2013) *Uncovering the politics of 'evidence' and 'results'. A framing paper for development practitioners*. www.bigpushforward.net. [accessed 2 August 2022].

Fetters, M.D., Molina-Azorin, J. (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.

Flybjerg, B. (2006) Making organization research matter: power, values and phronesis. In Stewart R. Clegg, Cynthia Hardy, Thomas B. Lawrence, and Walter R. Nord, eds., *The Sage Handbook of Organization Studies*, 2nd edition, Thousand Oaks, CA, Sage, pp. 370-387.

Garcia, R., Zazueta, A. (2015) Going beyond mixed methods to mixed approaches: a systems perspective for asking the right questions. *IDS Bulletin*, 46(1):1-14.

Gawande, A. (2008) *Better: a surgeon's notes on performance*. London: Profile.

Gibbs, A., Corboz, J., Chirwa, E. (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.

Glennerster, R., and Takavarasha, K. (2013) *Running randomized evaluations: a practical guide*. Princeton: Princeton University Press.

Goldstein, M., Pierotti, R. (2020) Mixing qualitative and quantitative methods. World Bank, *DevEval Blog*. https://blogs.worldbank.org/impactevaluations/mixing-qualitative-and-quantitative-methods-conversation [last accessed 17 August 2023]

Goertz, G. (2020) *Social science concepts and measurement*. Princeton University Press. 2nd Edition.

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

Guest, G., Fleming, P.J. (2015) Mixed methods research, in Guest, G. and Namey, E., editors, *Public Health Research Methods*. Thousand Oaks CA: Sage Publications.

Haig, B., Evers, C. (2016) *Realist inquiry in social science*. Los Angeles: Sage.

Harari, Y.N. (2011) Sapiens: *A brief history of humankind*. New York, N.Y: Harper.

Hayman, R., King, S., Kontinen, T., Narayanswarmy, L., editors (2016) *Negotiating knowledge: evidence and experience in development NGOs*. Practical Action Publishing: London.

Heinemann, E., Van Hemelrijck, A., Guijt, I. (2017) *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, July. ISBN 978-92-9072-767-5

Hernandez, K., Ramalingam, B., Wild, L. (2019) *Towards evidence-informed adaptive management: a roadmap for development and humanitarian organisations*. ODI working paper 565.

Holland, J., Campbell, J., editors (2005) *Methods in development research: combining qualitative and quantitative approaches*. ITDG Publishing: Rugby.

Honig, D., Pritchett, L (2019) *The limits of accountability in education (and far beyond): why more accounting will rarely solve accountability problems*. Washington, DC: Center for Global Development: working paper 510.

Howard, N. (2022) Towards ethical good practice in cash transfer trials and their evaluation. *Open Research Europe*. https://open-research-europe.ec.europa.eu/articles/2-12 [accessed 23 August 2023].

Humphreys, M., Jacobs, A. (2015) Mixing methods: a Bayesian approach. *American Political Science Review*, 109(4): 653-673.

Jayachandran, S., Biradavolu, M. and Cooper, J. (2023) Using machine learning and qualitative interviews to design a five-question survey module for women's agency, *World Development*, 161

Jimenez, E., Waddington, H., Goel, N., Prost, A., Pullin, A., White, H., Lahiri, S., Narain, A. (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):400-21.

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 and Capabilities*, 20:2, 197-217.

Margolies, A., Colantuoni, E., Morgan, R., Gelli, A., Caufield, L. (2023) The burdens of participation: a mixed-methods study of the effects of a nutrition sensitive agriculture program on women's time in Malawi. World Development, 162.

Martens, B., Mummert, U., Murrell, P., Seabright, P. (2002) *The institutional economics of foreign aid*. Cambridge: Cambridge University Press.

Maxwell, J. (2004) Using qualitative methods for causal explanation. *Field Methods*, 16. pp. 243–64.

Molecke, G., Pinkse, J. (2017) Accountability for social impact: a bricolage perspective on impact measurement in social enterprises. *Journal of Business Venturing*, 32(5):550-568.

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.

Morgan, J. (2019) A realist alternative to randomised control trials: a bridge not a barrier? *European Journal of Development Research*, 31(2):180-188.

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

Nicholls, A., Nicholls, J., Paton, R. (2015) Measuring social impact. Chapter 9 of Nichols, A., Paton, R., Emerson, J. editors, *Social Finance*. Oxford: Oxford University Press.

Neilsen, S., Jaspers, S., Lemire, S. (2023) The curious case of the realist trial: methodological oxymoron or unicorn? *Evaluation*, 1-18.

Pawson, R. (2013) *The science of evaluation: a realist manifesto*. Sage.

Picciotto, R., Weaving, R. (1994) A new project cycle for the World Bank. *Finance and Development*, December, 42.

Powell, S., Copestake, J., Remnant, F. (forthcoming) Causal mapping for evaluators. *Evaluation*.

Pritchett, L., Samji, S., Mammer, J. (2013) *It's all about MeE: learning in development projects through monitoring, experiential learning and impact evaluation*. Washington, DC: Centre for Global Development, working paper 233.

Ranganathan, M., Pichon, M., Hidrobo, M., Tambet, H., Sintayehu, W., Tadesse, S., Buller, A. (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.

Ravallion, M. (2018) *Should the randomistas (continue to) rule?* Washington, DC: Center for Global Development, working paper 492.

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., Walton, M. editors (2004) *Culture and public action*. Introduction. World Bank.

Rao, V., Walton, M. (2004) *Culture and public action*. World Bank and Stanford University Press.

Rodrik, D. (2008) *The new development economics: we shall experiment, but how shall we learn?* Harvard Kennedy School: John F. Kennedy School of Government, Research working paper 08-055.

Reichardt, C.S. (2022) The Counterfactual Definition of a Program Effect. *The American journal of evaluation*, 43(2), pp. 158–174.

Repko, A.F., Szostak, R. (2021) *Interdisciplinary research: process and theory*. Thousand Oaks CA: Sage Publications. 4th Edition.

Rogers, P. (2020). Real-Time Evaluation. *Monitoring and Evaluation for Adaptive Management. Working Paper Series*, No.4, December. www.betterevaluation [accessed 19 September 2023].

Ramalingam, B. (2013) *Aid on the edge of chaos*. Oxford: Oxford University Press.

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

Tashakkori, A. & Creswell, J. (2007) Editorial: the new era of mixed methods. *Journal of Mixed Methods Research*. 1(1):4-7.

Warren, E., Melendez-Torres, G., Bonell, C. (2022) Are realist randomised controlled trials possible? A reflection on the INCLUSIVE evaluation of a whole-school bullying prevention application. *Trials,* 23:82.

Webster, J., Exley, J., Copestake, J., Davies, R., Hargreaves, J. (2018) Timely evaluation in international development. *Journal of Development Effectiveness*, 10(4):482-508.

White, H. (2011) Achieving high-quality impact evaluation design through mixed methods: the case of infrastructure, *Journal of Development Effectiveness*, 3(1):131-144, DOI: 10.1080/19439342.2010.547588

White, H. (2019) Comment: the twenty-first century experimenting society: the four waves of the evidence revolution. *Palgrave Communications*, 5:47.

White, H., Raitzer, D.A. (2017) *Impact evaluation of development interventions: a practical guide*. Manila: Asian Development Bank.

White, S. (2015) Qualitative perspectives on the impact evaluation of girls' empowerment in Bangladesh. *Journal of Development Effectiveness*, 7(2):127-45.

Widner, J., Woolcock, M., Nieto, D.O. (2022) *The case for case studies: methods and applications in international development*. Cambridge University Press.

Woolcock, M. (2009) Towards a plurality of methods in project evaluation: a contextualised approach to understanding impact trajectories and efficacy. *Journal of Development Effectiveness*, 1:1-14.

Woolcock, M. (2019) *Reasons for using mixed methods in the evaluation of complex projects*. Center for International Development at Harvard University, Faculty Working Paper, No.348.