

Reflections from a Realist Evaluation in Progress: Scaling Ladders and Stitching Theory

Abstract Realist evaluation provides valuable insights into how and why programmes lead to change, and can generate transferable lessons to help practitioners roll out or scale up an intervention. However, as yet there are few standards and guidelines governing what counts as a ‘good’ realist evaluation. This CDI Practice Paper, written by Melanie Punton, Isabel Vogel and Rob Lloyd, reflects on the first year of a three-year realist impact evaluation, examining the Building Capacity to Use Research Evidence (BCURE) programme funded by the UK Department for International Development (DFID). It describes some of the challenges faced and lessons learned, providing insights into the potential value of realist approaches within international development.

When evaluating a programme’s impact it is important to understand *why* a particular change happened, and to investigate this in a robust way. Realist evaluation is a theory-based evaluation approach that goes beyond asking ‘what works?’ to investigate ‘how and why does this programme work or not work, for whom, and in what circumstances?’

The approach was developed 20 years ago (Pawson and Tilley 1997), but there are still few published examples, particularly within the international development field. While there is an active global community of practice engaging in enthusiastic methodological debates on the key principles of realist evaluation and how to apply them, there is currently an absence of standards governing what counts as a ‘good’ realist evaluation or guidance on the steps involved in conducting one.¹ This paper offers some reflections on the conceptual and practical lessons learned during the first year of an ongoing realist evaluation. We hope it will be of use to others thinking about or currently conducting realist evaluations within and beyond the field of international development.

1 What is realist evaluation?

Realist evaluation works by opening up the ‘black box’ between intervention and outcome, through developing and testing *programme theory*. A programme theory is

an explanation of how, why and in what contexts an intervention leads to particular outcomes. This explanation consists of linked sets of hypotheses about the *mechanisms* that cause an intervention to work or not work in particular *contexts*, to lead to specific *outcomes*. These hypotheses are known as ‘context–mechanism–outcome’ or CMO configurations (see Box 1) – the core analytical units of realist evaluation (Pawson and Tilley 1997; Wong *et al.* 2013).

Realist evaluation is underpinned by a ‘generative’ model of causality, in which causal links are demonstrated through a fine-grained explanation of what happens between cause and effect to explain *why* a certain effect occurs. Realist evaluation therefore explains programme impact in a different way to experimental and quasi-experimental evaluation approaches (which are underpinned by correlational or counterfactual models of causality) and those based on multiple-conjectural causal logic such as Qualitative Comparative Analysis (Schatz and Welle 2016; Befani 2012). Through developing and testing CMO configurations, realist evaluation demonstrates causality by providing an explanation of how and why a programme works.

Central to realist evaluation is a recognition of the fundamental importance of *context* in shaping how and why programmes lead to (or fail to lead to) change.

Box 1 Context–mechanism–outcome (CMO) configurations

Mechanisms are the causal forces, powers, processes or interactions that generate change within an intervention – including the choices, reasoning, and decisions that people make as a result of the resources provided by the programme. A training course is not a mechanism. The mechanism is the ‘thing’ that explains *why* training changes (or does not change) behaviour in a particular setting. For example, training may spark an ‘eye-opener’ for some participants, in which they recognise the relevance and value of the content in relation to their day-to-day work.

Mechanisms are only triggered in certain **contexts**. For example, an ‘eye-opener’ mechanism may not ‘spark’ if the trainees are doing jobs that will not allow them to put their new skills into practice.

Outcomes refer to intended and unintended short, medium and long-term changes resulting from an intervention.

A context–mechanism–outcome (CMO) configuration is a theory or hypothesis about how a particular mechanism works in a specific context to lead to an outcome. CMO configurations can be read as sentences, for example: ‘Where training content is directly relevant to a person’s day job (C), training on evidence-informed policymaking can spark an “eye-opener” in which trainees recognise how the principles can add value for them (M), leading to increased use of evidence in their day-to-day work (O).’

Source: Pawson and Tilley (1997); Westhorp (2014). Example from the Building Capacity to Use Research Evidence (BCURE) programme.

Programmes cannot simply be replicated in diverse contexts and expected to result in the same change, because contextual differences will cause the resources provided to trigger different mechanisms and lead to different outcomes. By providing insights into how and

why programmes work (or do not work) in different contexts, realist evaluation can help implementers learn how best to scale up or roll out a programme (Westhorp 2014). Box 2 explains the warrant for generalisability in realist evaluation.

Box 2 Realist evaluation and generalisability

The warrant for generalisability within realist evaluation derives from the realist philosophy underpinning the approach, which holds that causal mechanisms are *real* forces or processes that exist in the world and link causes to effects. This realist philosophy underpins the assumption that mechanisms are not unique to a particular setting. Rather, the same or similar mechanisms are present and explain causal links in different situations. If training can spark an ‘eye-opener’ that leads to behaviour change in one BCURE setting, the same mechanism may operate (given the right contextual conditions) to result in a similar outcome elsewhere.

CMO configurations are therefore ‘portable’. A tested set of CMO configurations from one programme can give practitioners a degree of confidence that an intervention might result in the desired change through the same mechanism elsewhere – when implemented in a certain way and where the contextual conditions are right. However, further investigation will be required to verify whether change does in fact happen in this way. Through further investigation, CMO configurations can be continually tested and refined, enabling the accumulation of knowledge about how programmes work in different contexts over time.

Source: Pawson and Tilley (1997); Westhorp (2014).

There is no straightforward ‘step-by-step’ framework for conducting realist evaluation. Westhorp (2014) describes it as ‘a way of thinking’ rather than a method. The approach can be successfully applied using a variety of qualitative and quantitative methods (depending on the issue under investigation) and the choice of methods will shape the specific steps taken in conducting the evaluation. However, realist evaluations always encompass three broad stages: *developing theory, testing theory, and refining theory*. These are iterative rather than linear; theory is developed, tested, refined and tested again as knowledge accumulates. Below, we introduce the BCURE evaluation and discuss how the evaluation team navigated each of these stages during its first year.

2 The Building Capacity to Use Research Evidence (BCURE) evaluation

Funded by the UK Department for International Development (DFID), BCURE works with policymakers in low-income and middle-income countries to develop skills, knowledge and systems in order to improve the use of evidence in decision-making. This £13m programme was launched in 2013 and is currently investing in a number of capacity development projects across Africa and Asia (see Box 3).

A realist approach was selected for the three-year BCURE impact evaluation because DFID was interested in understanding not just *whether* BCURE worked, but

how and why capacity building can contribute to increased use of evidence in policymaking. The primary aim of the evaluation was to strengthen the evidence base on how capacity building can promote evidence-informed policymaking, to inform decisions within and beyond DFID about whether to fund and how to design this type of programme in future.

The main components of the BCURE evaluation are as follows:

- 1 A realist literature review, synthesising published papers and grey literature related to capacity building for evidence-informed policymaking.
- 2 Six ‘programme evaluations’ of BCURE-funded projects, incorporating primary data collection and analysis of monitoring and implementation documents.
- 3 Additional primary research, aiming to generate evidence from *outside* the BCURE programme on how and why capacity building can result in better policies. This component is not discussed further in this paper.
- 4 A synthesis of findings from the above components, investigating how and why capacity building for evidence use works or does not work in different contexts.

Figure 1 illustrates how the components link to the three broad stages of a realist approach – developing, testing and refining theory. Data collection and synthesis is repeated each year for three years to enable the evaluation to track programme results over time, and iteratively test and refine theories about why these results have been achieved.

Box 3 About BCURE

The Building Capacity to Use Research Evidence (BCURE) programme consists of six separate projects, implemented by six different partners, across 11 countries in Africa and Asia. Each project works with government stakeholders with the aim of building capacity for evidence-informed policymaking.

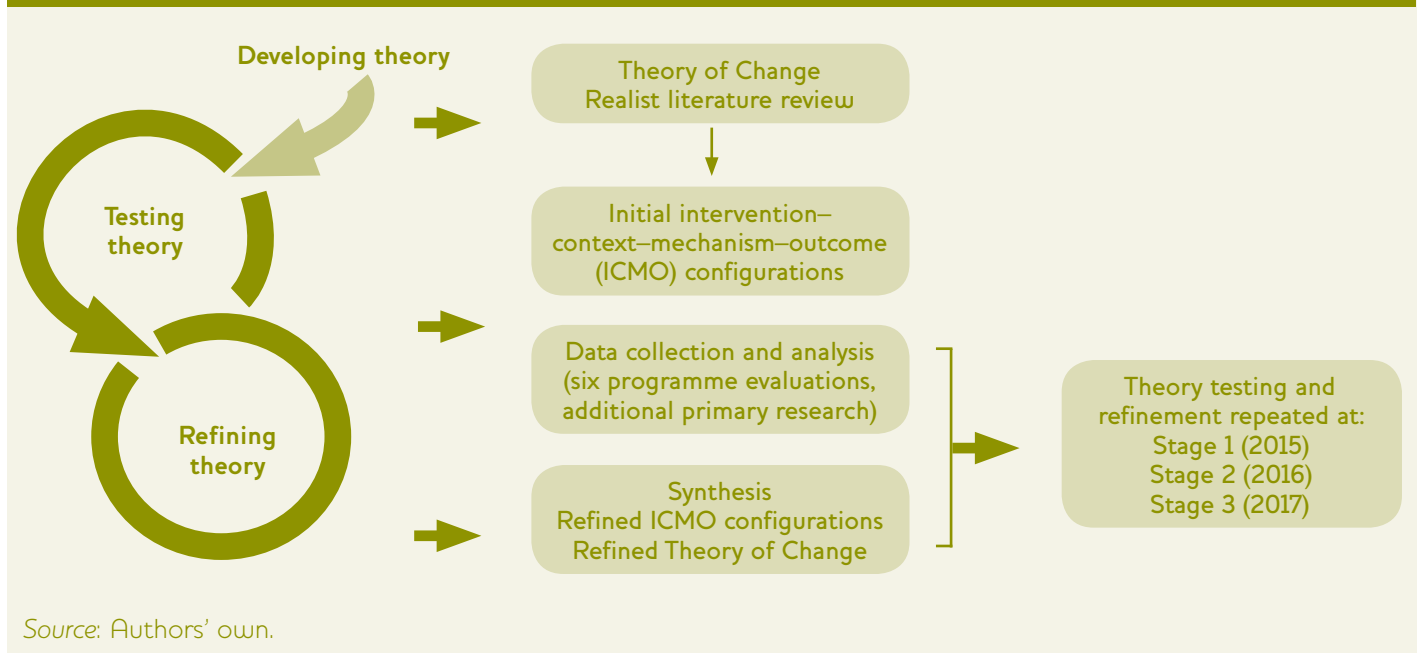
The projects involve a range of interventions, designed and combined in different ways by different partners. These include: training on how to access, appraise and use evidence in policymaking (online and face-to-face, in-workplace and residential, and shorter and longer in duration); practical workshops; mentoring; facilitating online and face-to-face networks; developing tools, systems and manuals to embed evidence use at an organisational level; and working with ‘evidence champions’ in government organisations.

Further information on BCURE is available at: www.itad.com/knowledge-and-resources/bcure.

3 Developing theory: opening up the ‘black box’ of capacity development for evidence-informed policy

The BCURE evaluation began by articulating an overarching theory of change (ToC) for the programme, depicting how and why we thought the various BCURE interventions might lead to improvements in capacity, changes in behaviour, and an increased use of evidence in policymaking.

Figure 1 The iterative stages of the BCURE evaluation



Source: Authors' own.

Box 4 The theory of change (ToC) and four levels of change

The literature on capacity building suggests that capacity change requires change throughout an organisational or institutional system, not just an accumulated change in individuals' skills and knowledge. We attempted to reflect this in the ToC by visualising four 'levels' of change, which the BCURE projects aim to influence, and which reinforce and feed into one another. The two-sentence version of the BCURE ToC is as follows:

'Developing the capacity of decision-makers to use research evidence (through building knowledge, skills, commitment, relationships and systems at individual, interpersonal, organisational and institutional levels) will allow them to access, appraise and apply good quality evidence more effectively when forming policy. This will improve the quality of policies, ultimately benefitting more poor people.'

Individual change includes individuals' skills and knowledge relating to searching for, appraising and

applying evidence in decision-making; as well as the motivation, attitudes, commitment and values that affect individual behaviour.

Interpersonal and network change refers to the relationships and networks between individuals and groups that influence evidence interpretation and use.

Organisational change refers to the systems, policies and procedures, practices, culture and norms within an organisation, which affect evidence access, appraisal and application in decision-making.

Institutional change relates to the wider environment within which individuals, networks and organisations operate. This includes the political system, civil society and the media, political and economic factors, donor influence, and broader social factors (culture, norms, collective beliefs) that influence the use of evidence in policy.

It is worth referencing the distinction between a realist 'programme theory' and a 'theory of change'. Blamey and Mackenzie (2007) suggest that 'programme theories' specifically articulate the hypothesised causal links in a programme, showing how mechanisms are expected to be triggered by an intervention to lead to anticipated outcomes. In contrast, they argue that 'theories of change' are more about the 'nuts and bolts' of a programme, mapping programme activities and their expected links to outcomes in the short, medium and long term.

Our BCURE ToC attempts to perform both functions. It spells out the anticipated links between activities and outcomes, but it also incorporates theory about *how and why* the interventions are expected to lead to change at different levels, and how these levels are expected to interact and reinforce one another (see Box 4). Our CMO configurations sit within the ToC and constitute more granular hypotheses about specific causal links and processes within it. We have used the term ToC rather than programme theory because it is a more widely used term in international development, and is more intuitive to BCURE partners.

The first iteration of the ToC drew on the evaluation team's existing knowledge (and professional hunches) about the nature of capacity building and the factors that affect evidence-use in policymaking. It also reflected the ToCs from the six BCURE projects.

The ToC was used to develop the research questions for a realist literature review.² We aimed to review studies of other interventions attempting to build capacity

for evidence-informed policymaking, in order to both refine our ToC and start developing CMO configurations. However, once we started to examine the concept of evidence-informed policymaking, the raft of assumptions underpinning the BCURE programme became increasingly apparent. What is 'evidence'? What makes it 'good quality'? What is 'policy' and what makes *that* good quality? What role does evidence play in policy processes and in making them 'better'? What does 'capacity' to conduct evidence-informed policymaking look like, and how can it be 'built'? This led to a deeper investigation of the theoretical literature than we originally intended – steering us into the waters of political science, psychology, adult learning theory and theories about complexity. These broader social science theories proved vital for fleshing out the theoretical frame for the evaluation.

Challenge 1: Developing, unravelling and re-stitching CMO configurations. We initially struggled to describe real-world situations in CMO configurations – a challenge that others have also noted (e.g. Dalkin *et al.* 2015; Adams *et al.* 2015). One issue is that the concept of 'mechanism' is widely used across the social sciences, but there is ambiguity in how it is applied by different researchers (Astbury and Leeuw 2010). We found ourselves asking repeatedly: is this really a mechanism, or is it a feature of the context affecting the mechanism, or a feature of the intervention itself?

Dalkin *et al.* (2015) suggested a modification to the CMO configuration in the attempt to clarify the concept of 'mechanism'. They suggest explicitly distinguishing between the *resource* introduced by a programme into

a specific context (e.g. a training course) and the change in *reasoning* this leads to (what happens in the heads of participants to lead to change). Along similar lines, within the BCURE evaluation, we decided to incorporate *features of the intervention* as an additional element to our CMO configurations. We felt this would help us to separate out features that are inherent in, or under the control of the programme (such as training design or length) from contextual factors that are not (such as professional incentives to participate in the training). Thus particular *features of an intervention* (I), introduced in a specific context (C), spark a *mechanism* (M) to lead to an *outcome* (O). This formula is referred to as an ICMO configuration for the remainder of this paper.

Our first ICMO configurations unravelled many times as our thinking developed, and had to be ‘stitched’ back together again. Considerable work and analytical effort is required to develop and refine this type of theory, which proved challenging given the time and resources available to us. We found two tactics very helpful when developing ICMO configurations. The first was to frame them as sentences, which can be intuitively understood without the need to continually grapple with the ICMO concepts. The second was to make use of metaphors and catchy names – a tip provided by a seasoned realist evaluator. These tactics proved helpful in facilitating conversations within the evaluation team, and when communicating findings to DFID and BCURE practitioners. Figure 2 depicts one of the BCURE evaluation ICMO configurations and the format used to present them.

Challenge 2: Deciding how many ICMO configurations to investigate. BCURE encompasses six projects, all using a different combination of interventions that work in different ways to lead to a variety of outcomes in the short, medium and long term. We struggled to decide on how many ICMO configurations were required to explain all of the various mechanisms within this complex programme.

Ultimately, we took a pragmatic route. We aimed to keep the number of ICMO configurations (and their degree of detail) at a level that could be reasonably investigated given available data and resources, and which seemed likely to provide the most operationally relevant insights – this was important, given that the main purpose of the BCURE evaluation is to inform decisions about future programming. For example, we decided against developing ICMO configurations to explain how BCURE interventions such as training result in increased skills and knowledge. This would require an investigation of psychological learning processes, which was not feasible within the scope of the evaluation. There is also widespread acceptance that developing skills and knowledge is not enough on its own to change behaviour – so ICMO configurations explaining how this happens are potentially less relevant to practitioners than those explaining how and in what contexts training leads individuals to use evidence more in their day-to-day work.

4 Testing theory: mastering the art of the realist interview and coding ICMO data

In a realist evaluation, decisions about sampling and data collection tools and methods are driven by consideration of who the researchers need to talk to in order to test their programme theory, and what the most appropriate tools and methods are for the job (Pawson and Tilley 1997).

Stage 1 data collection for BCURE took place in summer 2015. The evaluation team collected data in five countries – India, Zimbabwe, South Africa, Kenya and Sierra Leone. This involved up to 30 interviews per country with individuals who could provide an insight into how and why capacity building might influence evidence-informed policymaking (project staff, intervention participants, and informants from government and civil society). The team also reviewed programme materials and monitoring data for evidence on the nature of interventions and early outcomes.

Figure 2 A BCURE ICMO configuration

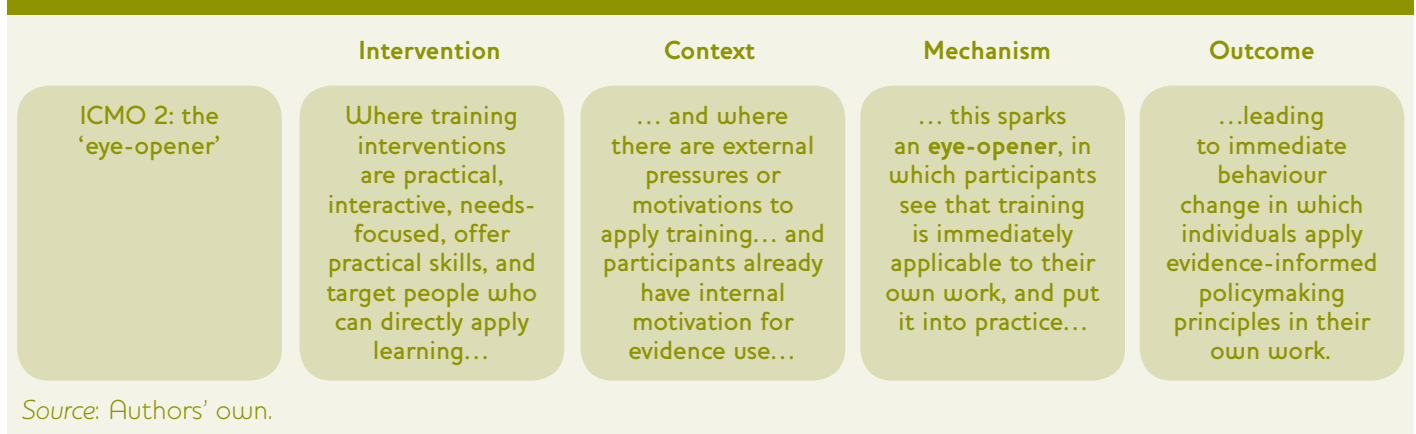


Table 1 Simplified version of the BCURE ICMO coding spreadsheet

Interviewee name / code	Country	Intervention type	Level mechanism is operating	Intervention factors	Contextual factors	Mechanism factors	Outcome	Outcome observed, anticipated or implied?
		Training	Individual	Training targets people at early stages of job: <i>'it was important that it happened as part of the induction process'</i>	Trainees <i>'don't have much to do yet'</i> so training came at the right time	Trainees developed good habits in using evidence as part of their day-to-day policy work	Trainees will use evidence in their day jobs going forward	Anticipated
		Guidelines to embed evidence in decision-making	Institutional	Guidelines developed through a participatory process involving high-level stakeholders	National health crisis enabled longer gestation period to gain cross-department buy-in	Revised processes facilitate Ministers and staff to do their work more efficiently, and are therefore valued	Evidence will be considered more systematically across departments when policy is developed	Anticipated

Source: Authors' own.

We attempted to collect data to develop and refine our initial ICMO configurations using a realist interviewing approach. The aim of a realist interview is to explicitly discuss the researchers' theory with the respondent (in this case, our ICMO configurations), giving him or her the chance to confirm, falsify and refine the theory (for example, 'This is what we think might be happening... What do you think?') (Pawson and Tilley 1997; Pawson 1996).

Challenge 3: Conducting a realist interview while avoiding confirmation bias. Our evaluation team had some reservations about conducting realist interviews. In some settings in Africa and Asia, it is considered rude or inappropriate to disagree with an 'expert'. Moreover, in an international development context where many governments are dependent on international funds for their programmes, respondents may be accustomed to giving encouraging responses to evaluators. In these cases, it may be difficult to explicitly test a theory without risking confirmation bias – the respondent agreeing with the researcher's theory for the sake of politeness or ensuring their favour.

At Stage 1, we therefore decided to avoid directly asking respondents for comments on our (still tentative) ICMO configurations. Rather, our interviewers asked respondents to consider how policymaking works in their setting.

Interviewers were asked to identify an example, and use that as the departure point to explore respondents' views on how and why the BCURE interventions might lead to change. The topic guides probed indirectly for reactions to our hypothesised ICMO configurations, and asked for examples to illustrate mechanisms in action.

Informal discussion with other realist evaluators has provided us with some ideas about how to approach the challenge of confirmation bias within a realist interview at Stage 2 of the evaluation – when it will be important to begin testing as well as refining our ICMO configurations. One option is to ensure that respondents are asked to provide concrete examples whenever they express agreement to a particular ICMO. Another is to ask respondents to adjudicate between rival ICMO configurations, rather than discussing one at a time.

Challenge 4: Analysing ICMO data. The team originally considered using qualitative analysis software to analyse interview data. However, we faced time and resource challenges, as well as the difficulty of multiple researchers in multiple countries needing access to and experience with the software. After some experimentation, we found that Microsoft Excel provided a simple solution to one of the potential pitfalls faced by realist evaluators – that of ensuring that ICMO data are considered in *configuration*

rather than atomised into separate context, mechanism and outcome factors (Pawson and Manzano-Santaella 2012). Creating a spreadsheet in Excel enabled us to 'code' data on ICMOs from each source on separate rows (see Table 1). Our coding was conducted as follows:

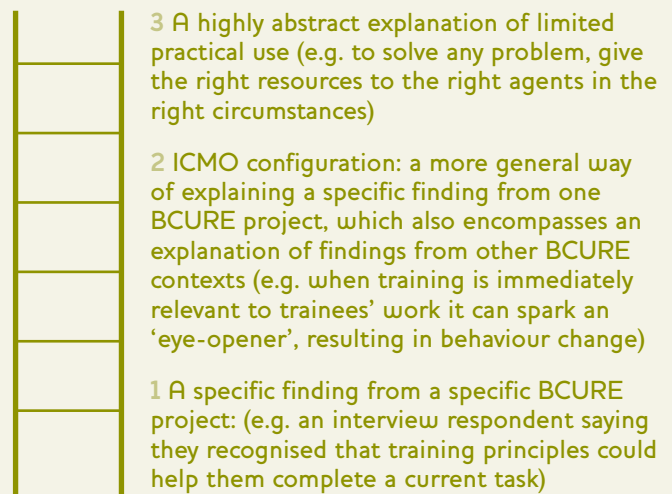
- We reviewed the interview transcripts for insights on how particular intervention features, implemented in a particular context, sparked such-and-such a mechanism – which led to (or is expected to lead to) a specific outcome. This process was an interpretive rather than mechanical one, requiring skill and judgement on the part of the researcher to decide how best to categorise the data.
- This information was entered (in summary form, along with verbatim quotes) in rows in an Excel spreadsheet.
- Where a source provided evidence of only *part* of an ICMO (for example, suggesting that a particular mechanism was important without providing any insights into the contextual or intervention factors that spark it), cells were simply left blank.
- However, the outcome cell was always completed, following the reasoning that any insight into relevant I, C or M factors must relate to an outcome. Each outcome was coded according to whether it was *observed* (the interview respondent stated that it had already happened), *anticipated* (it had not happened yet but the respondent expected it to), or *implied* (no explicit mention of the outcome was made but the interview data enabled the evaluation team to infer, tentatively, that the respondent had observed or anticipated it). In future stages of the evaluation, this coding approach will enable us to make judgements about the strength of evidence behind particular ICMOs.

This approach allowed us to easily search and filter our findings using various criteria, helping us to identify patterns in the data in order to refine our ICMO configurations. However, the process required considerable time and resources – more than originally anticipated. In addition, it was important for interview transcripts to be very detailed and include verbatim quotes to help identify insights into ICMO configurations.

5 Refining theory: climbing the ladder of abstraction

The data from across the six projects were then synthesised in order to draw out lessons and refine both the ICMO configurations and our ToC. We used a metaphor (the ladder of abstraction) and elements of a method (meta-ethnography) to help us do this in a clear, transparent and systematic way.

Figure 3 The ladder of abstraction



Source: Adapted from Cartwright and Hardie (2012).

A metaphor for refining theory: the ladder of abstraction

A 'ladder' metaphor is used to help conceptualise theory within realist evaluation (Westhorp 2012) as well as more broadly within and beyond the social sciences. The BCURE evaluation team drew on insights from Cartwright and Hardie (2012), who point out that 'there is always more than one correct way to describe what caused a result or justifies a prediction. Some of these ways will generalise across a great many cases, others across very few.' In realist evaluation terms, climbing the ladder of abstraction is about identifying findings from different BCURE contexts that point to the same underlying mechanisms (see Box 2). We synthesised data from across our cases through moving from specific findings at the bottom of the ladder (for example, a statement made by an interview respondent) up to more general explanations several rungs up (ICMOs), which encompassed findings from across different respondents and country settings.

A method for refining theory: meta-ethnography

We drew on meta-ethnography to provide a clear and transparent structure for the synthesis process. Meta-ethnography is an interpretive synthesis method, involving the transfer and translation of ideas, concepts and meanings across different sources (Noblit and Hare 1988). While we did not apply the method in full, we found two of its steps helpful to structure the synthesis: determining how the evidence was related, and 'translating' the sources into one another.

The synthesis process began with a two-day participatory evaluation team workshop, where the data were examined and the two steps applied.

Determining how the evidence was related. The team read through the coded ICMO data to identify two things.

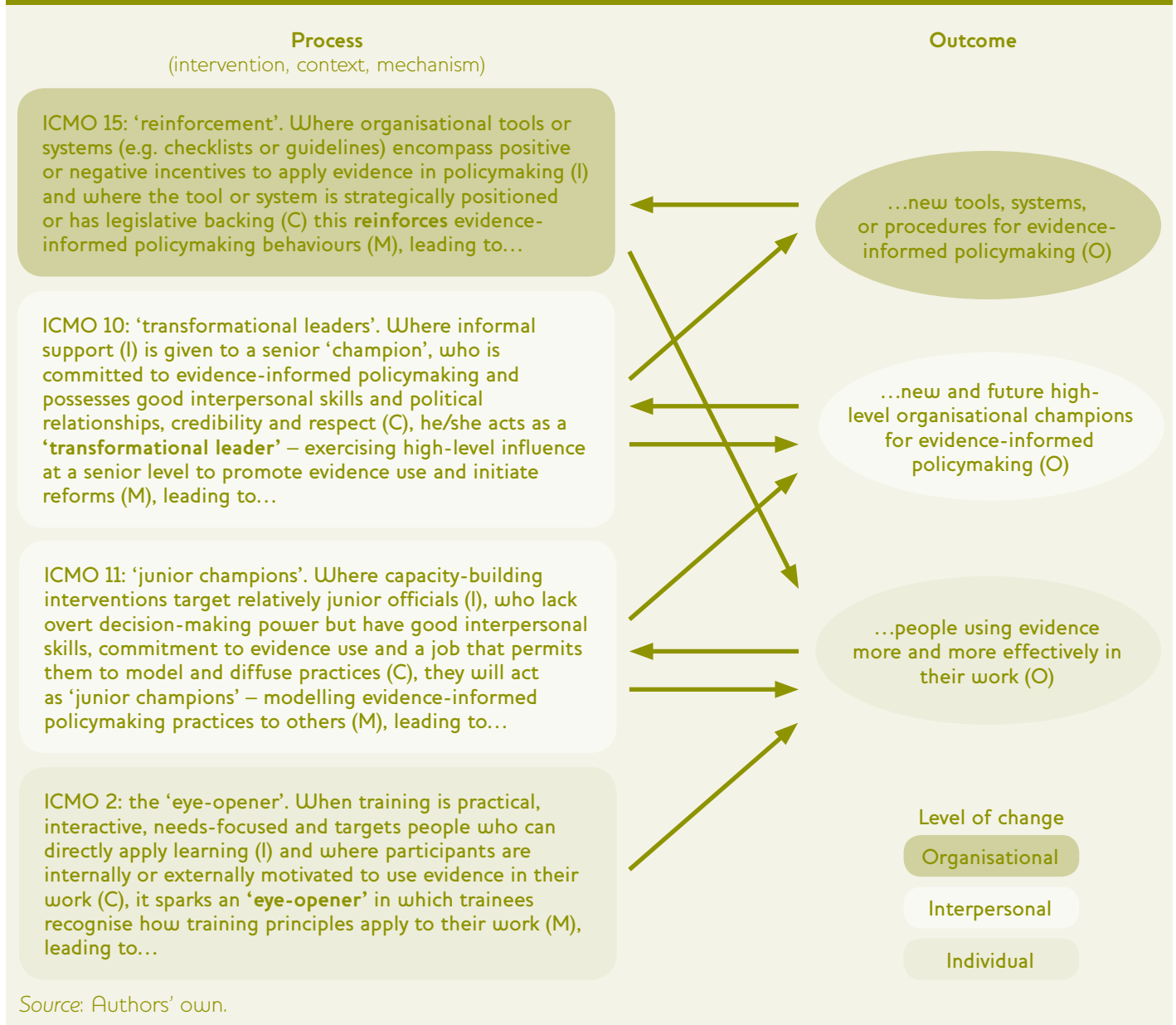
- 1 Common concepts, themes or metaphors which applied across the sources. In meta-ethnography these are known as *reciprocal translations*. These were identified by asking: 'Is this an example of something we have seen elsewhere? Is there a common concept we can use to explain these things?'
- 2 More abstract explanations or models that explained groupings of findings across the cases. These explanations are known in meta-ethnography as *lines of argument*. They exist further up the ladder of abstraction, and involve adding a new explanatory layer on top of the interpretations reached through reciprocal translation. Lines of argument were identified by asking: 'Can this concept, theme or metaphor be explained using a more *abstract* concept, theme or metaphor, which encompasses and goes beyond the more specific explanation?'

This analysis was used to start constructing new and revised ICMO configurations.

'Translating' new explanations across the cases: The emerging ICMO configurations were then *translated* across the original sources, through re-examining the data to consider how well they reflected and encompassed the ideas originally expressed in interviews. Team members were asked: 'Does this apply in your BCURE context? Are there any nuances from interviews with respondents in your setting?' This enabled scrutiny of differences within the data, which were used to adjust, refine and caveat our ICMO configurations (Pope, Mays and Popay 2007).

Following this exercise, two team members reviewed the full data set in a systematic way. We followed the example of other researchers who used tables, grids and

Figure 4 Interconnections between ICMOs at the individual, interpersonal and organisational levels



Source: Authors' own.

matrices when conducting meta-ethnographic synthesis (see, for example, Atkins *et al.* 2008; Britten *et al.* 2002). As well as the Excel database of ICMOs described above, tables in Word were used to help group data by ICMO, and reviewing these tables allowed us to further refine the ICMO configurations. Mid-way through this process, a BCURE workshop allowed us to share our thinking with the implementing partners for comment. This was treated as an additional translation step, allowing further identification of areas of agreement and disagreement, and refinements to the ICMO configurations. Throughout the synthesis, a record was kept of key analytical decisions to retain transparency about how theory was developed and refined.

At the end of the synthesis process, we had a revised set of ICMO configurations representing our 'best guesses' at the end of Stage 1 about how BCURE interventions are leading to change. These provided new insights into how elements of our ToC lead to and reinforce other elements. They were also used to refine our ToC by nuancing expected outcomes and adjusting the anticipated links between them. The results of our Stage 1 synthesis will be published in summer 2016, and our ICMOs and ToC will be revisited, tested and further refined at Stages 2 and 3 of the evaluation.³

Challenge 5: Encompassing complexity. Our final challenge relates to complexity. The BCURE literature review confirmed the importance of conceptualising capacity as a *complex* process, involving feedback loops, tipping points and non-linearity. The behaviours and interactions of participants in a capacity-building intervention are likely to combine in non-linear ways and amplify one another, resulting in feedback loops and emergent consequences that are not predictable in advance (Smith and Joyce 2012; Ramalingam 2013). Realist approaches are well-suited to evaluating complex issues; in fact, Westhorp (2014) argues that realism and complexity theory are 'natural bedfellows'. The recognition that programmes work through myriad mechanisms, which operate in different ways among different groups in different contexts, is an acknowledgement of this complexity. However, at times, the BCURE evaluation team felt shackled by our ICMOs, which are very much linear ('this intervention feature, in this context, sparks that mechanism to lead to this outcome' – see Figure 2).

We started to resolve this challenge during the Stage 1 synthesis process, drawing on work by Westhorp (2012) and Jagosh *et al.* (2015), which suggests that it is possible to conceptualise mechanisms in terms of 'levels of a system'. For example, a mechanism might be conceptualised at the level of an individual, a family, a community, or a society. An evaluator can 'layer' theories, with the *outcome* at one level becoming the *context* at the next level up, creating a ripple effect up the chain (Jagosh *et al.* 2015). Our data from Stage 1 suggest that ICMO configurations at different levels of our ToC (individual, interpersonal, organisational and

institutional) are interconnected in this way, with feedback loops operating at different levels (Figure 4).⁴ Attempting to conceptualise and measure this complexity has been one of our biggest challenges, but it is also one of the areas where we feel a realist approach can add the most value.

6 Conclusion

This paper has described the approach taken and challenges faced during the first year of a realist evaluation of the BCURE programme. We now reflect on some broader opportunities and challenges around applying a realist approach within an international development context.

Realist evaluation provides a systematic way of exploring context and complexity. There is now widespread recognition in the field of international development that understanding context and navigating complexity is crucial to successfully promote change. This recognition is reflected in the Doing Development Differently movement,⁵ and in ongoing conversations about 'adaptive programming' – the idea that programmes need to be ready and willing to adapt rapidly and flexibly in response to changing contexts. Realist evaluation provides a systematic (although certainly not easy!) way to examine how context affects the way people respond to the resources provided by programmes in a complex environment, and how this influences programme outcomes.

Realist evaluation generates operationally relevant findings that help improve and scale up programmes, although careful communication is needed when engaging commissioners and practitioners. Realist evaluation findings can provide a very practical steer on what types and features of interventions work best in a given context, as well as what practitioners need to think about when scaling up or rolling out a programme. We have found that the insights gained so far in the BCURE evaluation are intuitive and interesting to DFID and to the BCURE practitioners. However, there is an art in communicating realist evaluation findings without becoming mired in complex theoretical concepts such as CMO configurations. We have found certain strategies very useful in communicating our findings both within the evaluation team and outside it. Making use of metaphors has been invaluable, not only to help understand and navigate elements of the research process (e.g. scaling ladders, stitching theory) but also to help talk about our ICMO configurations (e.g. 'eye-openers'). It has also been very useful to frame ICMO configurations as sentences, largely avoiding the term 'ICMO' altogether when talking to practitioners (focusing instead on the content of the theory).

There are tensions between the theory-driven nature of realist evaluation and the structures and incentives of the international development sector. Given the iterative, theory-driven nature of realist evaluation, flexibility is

crucial. For example, the sample and methods may change over the course of the evaluation in order to test new and refined theory. However, evaluation commissioners' internal systems are not always conducive to flexibility – including procurement and programme management systems that require detailed up-front work plans and top-down budgeting and planning. Our experience reiterates the findings of others (Adams *et al.* 2015) in stressing the essential importance of engaging evaluation commissioners up front, and ensuring they are on board with the implications of a realist approach. The DFID team managing the BCURE evaluation are supportive in recognising the need for flexibility, which has been essential in enabling us to navigate the first year of the evaluation.

Overall, our experience from the first year of the BCURE evaluation suggests that the realist approach has much to offer the international development field. It provides a systematic way to examine how context affects

programme success within complex environments. It can also generate operationally relevant evidence on how and why change happens in different contexts, which can support the scale-up and roll-out of programmes. The BCURE evaluation team are hopeful that the evaluation will generate useful insights into what kinds of capacity development to promote evidence-informed policymaking lead to what sorts of outcomes, for whom, and in which contexts. However, developing, testing and refining theory requires considerable analytical investment, and the time and effort required should not be underestimated.

One major advantage of applying realist evaluation over a three-year period is the opportunity to refine and improve the approach as understanding evolves. Navigating a realist route through the first year of the BCURE evaluation has been challenging, but we look forward to continuing our realist journey and sharing our learning with the evaluation community.

Endnotes

- 1 The two-year RAMESES II project, launched in 2015, aims to develop quality and publication standards and training materials for realist evaluation. See <http://bmjopen.bmj.com/content/5/8/e008567>. Frequent discussion among the realist community of practice takes place through the Realist and Meta-narrative Evidence Synthesis: Evolving Standards mailing list. See www.jiscmail.ac.uk/.
- 2 The literature review is available from www.itad.com/knowledge-and-resources/bcure.

- 3 The synthesis report will be available here: www.itad.com/knowledge-and-resources/bcure.
- 4 Although we have a large amount of data on institutional-level factors and how they affect evidence use in policy processes, we have not yet worked through the best way to conceptualise institutional-level ICMOs and their interlinkages with other levels of the system. This will be a focus of Stage 2 data collection and synthesis.
- 5 <http://doingdevelopmentdifferently.com/>.

References

- Adams, A.; Sedalia, S.; McNab, S. and Sarker, M. (2015) 'Lessons Learned in Using Realist Evaluation to Assess Maternal and Newborn Health Programming in Rural Bangladesh', *Health Policy and Planning*, www.heapol.oxfordjournals.org/cgi/doi/10.1093/heapol/czv053 (accessed 16 March 2016)
- Astbury, B. and Leeuw, F.L. (2010) 'Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation', *American Journal of Evaluation* 31.3: 363–81, <http://aje.sagepub.com/cgi/doi/10.1177/1098214010371972> (accessed 21 March 2014)
- Atkins, S.; Lewin, S.; Smith, H.; Engel, M.; Fretheim, A. and Volmink, J. (2008) 'Conducting a Meta-ethnography of Qualitative Literature: Lessons Learnt', *BMC Medical Research Methodology* 8.1: 21, www.biomedcentral.com/1471-2288/8/21 (accessed 16 March 2016)
- Befani, B. (2012) 'Models of Causality and Causal Inference', Background paper to DFID Working Paper 38, *Broadening the Range of Designs and Methods for Impact Evaluations*, London: Department for International Development
- Blamey, A. and Mackenzie, M. (2007) 'Theories of Change and Realistic Evaluation: Peas in a Pod or Apples and Oranges?', *Evaluation* 13.4: 439–55, <http://evi.sagepub.com/content/13/4/439.full.pdf+html> (accessed 31 March 2014)
- Britten, N.; Campbell, R.; Pope, C.; Donovan, J.; Morgan, M. and Pill, R. (2002) 'Using Meta Ethnography to Synthesise Qualitative Research: A Worked Example', *Journal of Health Services Research & Policy* 7.4: 209–15, www.ncbi.nlm.nih.gov/pubmed/12425780 (accessed 16 March 2016)
- Cartwright, N. and Hardie, J. (2012) *Evidence-Based Policy: A Practical Guide to Doing It Better*, New York: Oxford University Press
- Dalkin, S.M.; Greenhalgh, J.; Jones, D.; Cunningham, B. and Lhussier, M. (2015) 'What's in a Mechanism? Development of a Key Concept in Realist Evaluation', *Implementation Science* 10.1: 49, www.implementationscience.com/content/10/1/49 (accessed 21 April 2015)
- Jagosh, J.; Bush, P.; Salsberg, J.; Macaulay, A.; Greenhalgh, T.; Wong, G.; Cargo, M.; Green, L.W.; Herbert, C.P. and Pluye, P. (2015) 'A Realist Evaluation of Community-based Participatory Research: Partnership Synergy, Trust Building and Related Ripple Effects', *BMC Public Health* 15.1: 725, www.biomedcentral.com/1471-2458/15/725 (accessed 16 March 2016)
- Noblit, G. and Hare, R. (1988) *Meta-Ethnography: Synthesizing Qualitative Studies*, California: Sage Publications

- Pawson, R. (1996) 'Theorizing the Interview', *The British Journal of Sociology* 47.2: 295–314
- Pawson, R. and Manzano-Santaella, A. (2012) 'A Realist Diagnostic Workshop', *Evaluation* 18.2: 176–91
- Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*, London: SAGE
- Pope, C.; Mays, N. and Popay, J. (2007) *Synthesizing Qualitative and Quantitative Health Evidence: A Guide to Methods*, Maidenhead: McGraw-Hill International
- Ramalingam, B. (2013) *Aid on the Edge of Chaos: Rethinking International Cooperation in a Complex World*, New York: Oxford University Press
- Schatz, F. and Welle, K. (2016) *Qualitative Comparative Analysis: A Valuable Approach to Add to the Evaluator's Toolbox? Lessons from Recent Applications*, CDI Practice Paper 13, www.ids.ac.uk/publication/qualitative-comparative-analysis-a-valuable-approach-to-add-to-the-evaluator-s-toolbox-lessons-from-recent-applications (accessed 16 March 2016)
- Smith, K.E. and Joyce, K.E. (2012) 'Capturing Complex Realities: Understanding Efforts to Achieve Evidence-based Policy and Practice in Public Health', *Evidence & Policy: A Journal of Research, Debate and Practice* 8.1: 57–8, <http://openurl.ingenta.com/content/xref?genre=article&issn=1744-2648&volume=8&issue=1&spage=57> (accessed 16 March 2016)
- Westhorp, G. (2014) *Realist Impact Evaluation: An Introduction*, London: Overseas Development Institute/Methods Lab, www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9138.pdf (accessed 16 March 2016)
- Westhorp, G. (2012) 'Using Complexity-consistent Theory for Evaluating Complex Systems', *Evaluation* 18.4: 405–20
- Wong, G.; Westhorp, G.; Pawson, R. and Greenhalgh, T. (2013) *Realist Synthesis: Rameses Training Materials*, London: National Institute for Health Research (NIHR) Health Services and Delivery Research (HS&DR) Programme, www.ramesesproject.org/media/Realist_reviews_training_materials.pdf (accessed 16 March 2016)

“ Our experience from the first year of the BCURE evaluation suggests that the realist approach has much to offer the international development field. It provides a systematic way to examine how context affects programme success within complex environments. It can also generate operationally relevant evidence on how and why change happens in different contexts, which can support the scale-up and roll-out of programmes. ”

Centre for Development Impact (CDI)

The Centre is a collaboration between IDS (www.ids.ac.uk), Itad (www.itad.com) and the University of East Anglia (www.uea.ac.uk).

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

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

This CDI Practice Paper was written by **Melanie Punton**, **Isabel Vogel** and **Rob Lloyd**.

The opinions expressed are those of the author and do not necessarily reflect the views of IDS or any of the institutions involved. Readers are encouraged to quote and reproduce material from issues of CDI Practice Papers in their own publication. In return, IDS requests due acknowledgement and quotes to be referenced as above.

© Institute of Development Studies, 2016
ISSN: 2053-0536
AG Level 2 Output ID: 623

