METHODS LAB

September 2014

REALIST IMPACT EVALUATION

AN INTRODUCTION

Gill Westhorp

KEY MESSAGES

- Realist evaluation is a member of a family of theory-based evaluation approaches which begin by clarifying the 'programme theory': the mechanisms that are likely to operate, the contexts in which they might operate and the outcomes that will be observed if they operate as expected.
- Realist approaches assume that nothing works everywhere for everyone: context makes a big difference to programme outcomes. A realist evaluation asks not 'what works?' but 'how or why does this work, for whom, in what circumstances?'
- Realist impact evaluation is most appropriate for evaluating new initiatives or programmes that seem to work but where 'how and for whom' is not yet understood; programmes that have previously demonstrated mixed patterns of outcomes; and those that will be scaled up, to understand how to adapt the intervention to new contexts.

A METHODS LAB PUBLICATION

ODI.ORG/METHODSLAB

Australian Government

Department of Foreign Affairs and Trade

The Methods Lab is an action-learning collaboration between the Overseas Development Institute (ODI), the Australian Department of Foreign Affairs and Trade (DFAT) and BetterEvaluation. The Methods Lab seeks to develop, test, and institutionalise flexible approaches to impact evaluations. It focuses on interventions which are harder to evaluate because of their diversity and complexity or where traditional impact evaluation approaches may not be feasible, with the broader aim of identifying lessons with wider application potential.

Overseas Development Institute
203 Blackfriars Road
London SE1 8NJ

Tel. +44 (0) 20 7922 0300
Fax. +44 (0) 20 7922 0399
E-mail: info@odi.org.uk

www.odi.org
www.odi.org/facebook
www.odi.org/twitter

BetterEvaluation
E-mail: bettereval@gmail.com
www.betterevaluation.org

Introduction

This introduction has been developed to help evaluators and commissioners of evaluation to decide whether a realist impact evaluation is appropriate and feasible for a particular policy or programme. The term 'impact evaluation' is used here to refer to evaluations which examine the direct and indirect contribution of an intervention to changes in people's lives, especially longer-term changes.

This paper is product of the Methods Lab, an action-learning collaboration between the Overseas Development Institute (ODI), the Australian Department of Foreign Affairs and Trade (DFAT) and BetterEvaluation. The Methods Lab seeks to develop, test, and institutionalise flexible approaches to impact evaluations. It focuses on interventions which are harder to evaluate because of their diversity and complexity or where traditional impact evaluation approaches may not be feasible, with the broader aim of identifying lessons with wider application potential.

The paper provides a very brief introduction to the key ideas in realist evaluation and the implications of a realist approach for impact evaluation questions, evaluation design, programme theory, data collection, analysis and utilisation. It discusses:

- the benefits and challenges of a realist approach
- the features of programmes, policies and evaluations that lend themselves to a realist approach
- factors to take into account in undertaking evaluability assessment for a realist evaluation
- features of good practice in realist evaluation.

Illustrations are drawn from a number of projects that are, at the time of writing, under consideration for impact evaluation by the Methods Lab. References and further reading are listed at the conclusion of the paper.

A note on terminology: impact and outcome

In 2011, the DFAT, released a discussion paper on impact evaluation for practitioners. Box 1 gives the paper's definition of impact evaluation.

Realist evaluation uses the term 'outcome' to include short, medium and long term changes, intended and unintended, resulting from an intervention. The only difference between the terms 'impact' (as defined above) and 'outcome' (as used in realist evaluation) is that 'impact' implies changes "for people and their lives"; whereas 'outcome' includes change for people and their lives but can also include other kinds of changes (for organisations, workers, governments and so on).

The term 'outcome' is used in this paper for consistency with realist evaluation conventions.

Box 1. DFAT Definition of Impact (AusAid 2012)

Impacts are positive or negative changes produced by a development intervention – directly or indirectly, intended or unintended – in the context of its environment, as it interacts with the multiple factors affecting development change.

Impact occurs at multiple levels and timeframes – there can be short-term, intermediate and long-term changes resulting from an intervention. How and when impact occurs will differ depending on the type of intervention and the context.

An impact evaluation is a systematic and empirical investigation of the impacts produced by an intervention - specifically, it seeks to establish whether an intervention has made a difference in the lives of people. It aims to answer questions about what works or does not work, how, for whom, and why.

To provide these answers, impact evaluation links cause and effect: it assesses the direct and indirect causal contribution of the intervention to change in people's lives.

Therefore, an impact evaluation:

- evaluates the positive and negative, primary and secondary long-term impacts that result from a development intervention
- assesses the direct and indirect contribution of these interventions to such impacts on final beneficiaries, especially for the poor, whether intended or unintended
- explains how policy interventions contribute to an impact so that lessons can be learned. (p2)

What is realist evaluation?

The term 'realist evaluation' is drawn from Pawson and Tilley's seminal work, *Realistic Evaluation* (1997)[1]. It is a member of the family of theory-based evaluation approaches. Theory-based evaluation starts by clarifying the 'programme theory' – that is, clarifying how programme activities are understood to cause (or contribute to) outcomes and impacts. What distinguishes realist evaluation from other forms of theory-based evaluation are the particular assumptions that realist philosophy[2] makes about the nature of reality, how causation works, and what these assumptions imply for evaluation design, methods and utilisation.

Realist approaches assume that nothing works everywhere or for everyone, and that context really does make a difference to programme outcomes. Consequently, policy-makers and practitioners need to understand how and why programmes work and don't work in different contexts, so that they are better equipped to make decisions about which programmes or policies to use and how to adapt them to local contexts.

Consequently, realist evaluation does not ask 'what works?', 'does this work?' or (retrospectively) 'did this work this time?'. A realist research question contains some or all of the elements of "how and why does this work and/or not work, for whom, to what extent, in what respects, in what circumstances and over what duration?'.

Causation (how programmes cause change) and attribution (whether observed changes can be attributed to the programme or were caused by other things) are critical questions for impact evaluation. If an evaluation does not address these questions, it is not an impact evaluation. If it does not use a realist understanding of causation to do so, it is not a realist impact evaluation.

Five key ideas in realism and their implications for evaluation

Realism is a school of philosophy. It was developed to sit between positivism ('there is such a thing as the real world, which we can directly observe and about which we can derive "facts"') and constructivism ('since all our observations are shaped and filtered through human senses and the human brain, it is not possible to know for certain what the nature of reality is'). All evaluation approaches – consciously or unconsciously – reflect deep philosophical assumptions. Here we summarise, in very brief form, five key ideas from realism and their implications for evaluation, including impact evaluation.

1. ***Realism asserts that both the material and the social worlds are 'real', at least in the sense that anything that can have real effects is itself real.*** Thus, for example: gender - as distinct from sex - is real (and we know that it is because it has real effects); culture is real; class is real and so on.

   This has two main implications for evaluation. Firstly, it implies that programmes and policies are also 'real' and can have real effects – positive and negative, intended and unintended. Secondly, it implies that social institutions and constructs (culture, class, gender, religion, political and economic systems...) will have real effects on whether and how programmes work.
2. ***Realism acknowledges that all enquiry and observation are shaped and filtered through the human brain and that there is, therefore, no such thing as 'final' truth or knowledge.*** Nonetheless it argues that it is possible to work towards a closer understanding of the nature of reality. This is because reality itself constrains the interpretations that can reasonably be made of it. (For example: it is more reasonable to believe that I am, at this moment, sitting writing at a computer than hiking in the Himalayas.)

   The implication for evaluation is that it is possible to work towards better understanding of whether, how and why programmes work, even though we can never reach final certainty or provide definitive 'proof'.
3. ***Realism argues that all social systems are open systems.*** Their boundaries are porous and flexible: people, ideas, information and resources flow in and out of social systems. Social systems themselves interact and influence each other. Families and schools; the economic system and the political system – each interacts with, affects and is affected by the other.

   This has three implications for evaluation. Firstly, programmes themselves are open social systems. Evaluating them requires at least a general understanding of systems theory, an understanding of appropriate methods for systems evaluation and an understanding of the specific programme system. Secondly, it is necessary to choose the boundaries of the system(s) that will be included in the evaluation, even though the boundaries do not exist in this clear way in reality. Thirdly, those systems will not be static, but will change over time, in complex and interactive ways – regardless of whether a programme or policy is introduced.

   Perhaps most importantly: any outcome that is observed will be a result of interactions within and across systems – not simply an outcome of the programme

1 Pawson, R. and Tilley, N. (1997) *Realistic Evaluation,* Sage.

2 There are many schools of realism. Pawson and Tilley call their approach 'scientific realism'.

or policy. Any outcome will be a result of many causes and any action or change in a system may have many consequences. If causation itself is not simple and linear, evaluation approaches that treat causation as being simple and linear will be inadequate. This is perhaps particularly important in impact evaluations because they often investigate quite long 'causal chains' (from policy or funding initiative to final outcomes for people) and over sometimes quite long time periods. The longer the causal chain or time period, the more interactions there will be.

4. ***Realism offers a particular understanding of how causation works.*** The basic idea is that things that we experience or can observe are caused by 'deeper', usually non-observable processes. So, for example: we can open our hand and observe the tennis ball we held fall to the ground, but we cannot 'see' the gravity that causes the ball to fall. Similarly, we can experience a training programme and observe that participants use different language at the end of it than they did at the beginning, but we cannot 'see' the new content being stored in memory or the new connections being forged in the brain that enables them to do so. That is, the causal processes happen at a different level of the system than the observable outcomes. In realist philosophy, the underlying causal process is known as a 'mechanism'.

There are two other important things to understand about the idea of mechanisms. The first is that they exist as part of a whole system. A trainer only has the power to 'cause' change because he or she operates in relation to a student (or group of students), in a training programme, using the spaces, equipment and materials provided, and drawing on the social rules that guide teaching and learning. If any of these elements of the system are removed or changed, the causal process changes too. The second is that mechanisms exist whether or not they are operating at a particular moment. Gravity exists whether or not I let go of the tennis ball. The trainer has the power to teach, and the learner has the power to learn, whether or not they are currently doing so. The mechanism is already there, but it will only operate when the circumstances are right.

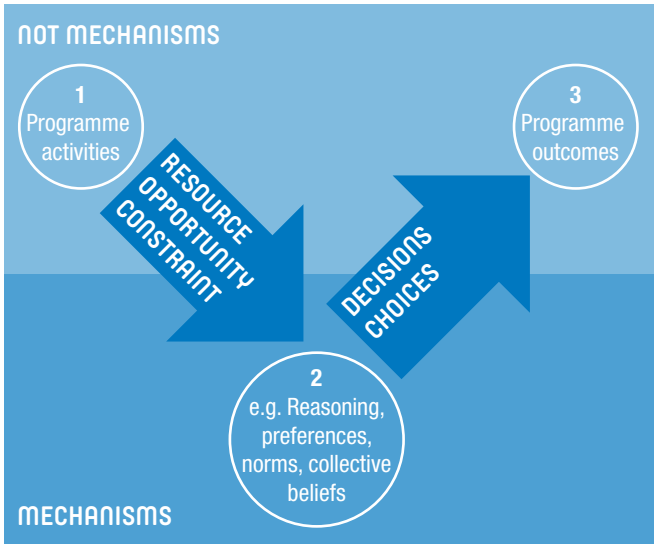
Pawson and Tilley (1997) were the first to clarify the implications of the realist understanding of 'mechanisms' for programme and policy evaluation. They asked what the 'causal powers' of programmes might be. They argued that programmes provide something – a resource, an opportunity or a constraint of some kind – *that is intended to influence the target person's decision-making*. Social policies often aim to shift the proportion of a population that can or will make a desired decision (e.g. boiling water before drinking it, sending children to school, growing more productive crops in more environmentally sustainable ways, and so on). However, ultimately, it is the target person's decision that determines whether the desired outcome is achieved (e.g. reduction in water-borne diseases, improved education outcomes, greater food security). That is, it is *the interaction between what the programme provides and the reasoning of its intended target population* that causes the outcomes. This interaction, therefore, constitutes a 'programme mechanism'.[3] The short-hand for this in realist circles is 'reasoning and resources'. The implication is that the evaluator needs to identify what resources, opportunities or constraints were in fact provided, and to whom; and what 'reasoning' was prompted in response, generating what changes in behaviour, which in turn generate what outcomes.

It is critical to understand the idea of 'mechanisms' for realist impact evaluation.

First, mechanisms cannot be directly observed simply by observing what programmes do. It is necessary to know what occurred because this provides information about the resources that were provided, but this does not tell us about the responses of intended targets. Information about those responses – who responds in what ways and why - is needed to understand mechanisms.

Second, mechanisms are not usually visible but they can be - and in a realist evaluation, are - investigated. It is possible to investigate gravity and how it works; or learning and how it works – but (to return to our earlier examples) doing so requires different research

Generative explanation in realist programme evaluation



Wong, G., Westhorp, G., Pawson R, and Greenhalgh, T. (2012) Realist Synthesis RAMESES Training Materials. Reproduced with permission. Note that 'generative explanation' means 'explaining how causation works'.

3 This is only one way of understanding the idea of 'mechanism'.

procedures than observing the tennis ball falling or the student in the training programme talking.

In realist evaluation, ideally, the evaluator hypothesises in advance the mechanisms that are likely to operate, the contexts in which they might operate and the outcomes that will be observed if they operate as expected. This is known as developing Context-Mechanism-Outcome (CMO) hypotheses. For example, providing free tuition might encourage poor parents to send their children to school, contributing to better education outcomes. However, it might also make parents feel as though they don't have the right to question what is provided by the school because they are not paying for it, which might contribute to lower community engagement with the school – which undermines education outcomes. It might also influence wealthier parents to send their children to private schools, resulting in greater segregation between poorer and richer children – again undermining education outcomes. In order to find out what impacts free tuition has, data about BOTH the mechanisms and the outcomes will have to be collected as part of the evaluation. If the evaluation only collects data about outcomes, it will not be possible to identify what caused the outcomes. If this happens, policy and programme staff will not know how to replicate the outcomes in another setting, because they will not know how they were caused in the first case. If the evaluation only collects data about expected mechanisms, it will not be possible to say whether the anticipated outcomes were achieved.

Third, the 'reasoning' of intended beneficiaries is socially and culturally conditioned. Men and women, adults and children, people of different cultural groups, religions, classes or economic situations may respond quite differently to exactly the same programme resource, such as training or a loan. Their values and beliefs, norms, cultural roles, previous experiences, and current circumstances will shape their reasoning, their decisions and, therefore, programme outcomes. This leads to patterns of outcomes that can be observed in most programmes. Mechanisms can be intended and unintended; some generating positive outcomes and some generating negative outcomes. The trick for realist evaluation is to work at a useful level of abstraction and to consider the *main* mechanisms generating the *main* patterns of outcomes.

Fourth, programmes may work by *enabling existing* reasoning ("I'd like to do this but I can't because I don't have the resources to do so – this programme provides the resources to do so") or by *changing* reasoning ("I have a new understanding of the value or importance of 'x' so I will now do 'y'"). It is important to remember here that 'reasoning' is a catch-all term for 'anything that happens inside the intended beneficiary's head': it can imply reasoning in the sense of logic-in-use, or values, emotion, or any combination of these.

Finally, mechanisms do not only operate at the end point of a programme with the target population. Most programmes have long implementation chains with different decisions being made by different levels of government, or organisations, or levels within organisations, along the way. This means that mechanisms are operating (or not, depending on the context) at all stages along a programme implementation chain. The implication for evaluation is that it is necessary to decide the points along the implementation chain that will be examined within the evaluation, and to hypothesise the main mechanisms and their outcomes for those points – not necessarily just the final point.

Programmes can generate outcomes at many levels of a system – individual community members, organisations, service delivery systems and so on. Where outcomes are to be investigated at multiple levels, understanding of context, mechanism and outcome will be required at multiple levels.

5. ***Realism provides a specific way of thinking about 'context'.*** Whether mechanisms 'fire' (operate at a particular moment) depends on the context. If I am standing on land when I release the tennis ball, gravity will draw it to the ground. If I happen to be underwater, a different mechanism (buoyancy) will cause the ball to float. In fact both mechanisms operate in both contexts, but gravity is the stronger force in air and buoyancy the stronger force under water. The two mechanisms compete, and the context determines which mechanism 'wins'.

   In the example of free tuition discussed above, different types of parents responded to the same resource in different ways, generating different outcomes. Here, the different sub-groups of parents were the important feature of context. This is known in realist evaluation as a Context-Mechanism-Outcome configuration (or pattern), often abbreviated to CMO or CMOC.

   The implication for evaluation is that what matters about context is what influences <u>whether</u> mechanisms operate, and <u>which</u> mechanisms operate.

   Context can influence programme mechanisms in many different ways. The context within which the organisation implementing a programme can influence the way in which, or the extent to which, a programme is implemented, who it targets, who it reaches and so on. However, it can also influence the ways in which intended beneficiaries respond. Responses may differ depending on whether the programme is delivered by government or non-government agencies, for example, if trust in one sector is lower than for the other. Variations within the target population can influence which mechanisms operate (gender, class, caste, culture and so on), which is the basis of the "for whom" question in realist evaluation. Access to resources to implement and opportunities to implement decisions, can also influence reasoning itself, as well as whether or not desired choices can be put into action.

A realist evaluation therefore hypothesises which features of context are likely to affect how, and for whom, a programme is expected to work, and collects data about those features of context. It then also needs data and analytic strategies to examine the interaction between context and mechanism.

These five basic understandings about the nature of reality, what we can know about reality, and how programmes work have implications for policy and practice and, therefore, for impact evaluation. Because programmes work differently in different contexts and through different change mechanisms, we cannot assume that programmes can be replicated from one context to another or that they will automatically achieve the same outcomes if they are. What is portable, however, are good understandings about 'what works for whom, in what contexts, and how'.

Therefore, realist impact evaluation does not simply try to ascertain the impacts of a particular programme in a particular place at a particular time. Realist impact evaluation seeks to inform policy and practice by learning more about 'what works for whom', 'in which contexts particular programmes do and don't work', and 'what mechanisms are triggered by what programmes in what contexts'. These understandings can then inform choices about which programmes to trial in what contexts, how to refine policies and programmes to improve their effectiveness, and how to adapt programmes to new contexts.

"In summary, realism holds that mechanisms matter because they generate outcomes, and that context matters because it changes... the processes by which an intervention produces an outcome. Both context and mechanism must therefore be systematically researched along with intervention and outcome. By implication, research or evaluation designs that strip away or 'control for' context with a view to exposing the 'pure' effect of the intervention limit our ability to understand how, when and for whom the intervention will be effective. " (Wong et al 2013, p. 13)

In what circumstances is realist impact evaluation suitable?

A realist approach is not a method but a way of thinking, so realist design can be incorporated within almost any impact evaluation. However, there are some purposes for which realist evaluation is particularly suited, some circumstances that make a realist approach more appropriate than other methods, and some circumstances in which a realist evaluation is easier than other circumstances.

For what purposes is realist impact evaluation most appropriate?

Realist evaluation is designed to improve understanding of how and why interventions work or do not work in particular contexts. It has a particular focus on understanding causation and understanding why different outcomes are achieved in different contexts. It is, therefore, particularly appropriate:

- for evaluating new initiatives, pilot programmes and trials, or programmes that seem to work but 'for whom and how' is not yet understood;
- for evaluating programmes that will be scaled out, to understand how to adapt the intervention to new contexts;
- for evaluating programmes that have previously demonstrated mixed patterns of outcomes, to understand how and why the differences occur.

Realist evaluation can help inform decisions at different levels related to the intervention, see Table 1 (p.8) for a few examples.

However, no evaluation approach is suited to all purposes or situations. Realist evaluation is not appropriate when:

- how, why and where programmes work is already well understood – a realist evaluation is not required and monitoring of implementation and outcomes should be sufficient;
- there is no real interest in understanding how the programme works – development of programme theory and theory-based data collection are unlikely to be supported;

Table 1. Benefits of a realist approach for decision-making at different levels

Role	Function	Benefits of a realist approach
Programme staff	Implement programmes	Detailed understanding of CMO configurations enables tailoring of programmes to local contexts Tailoring of programmes improves effectiveness
Policy staff	Select or design and administer programmes	Understand which programmes, from a suite of possibilities, are best suited to particular contexts Improve quality of advice to senior decision-makers Tailor programme design or guidelines to national or state contexts
Politicians Funding bodies	Fund programmes	Understand why a suite of programmes may be required to meet particular policy objectives Understand which programmes, from a suite of possibilities, are best suited to particular contexts Understanding programme mechanisms may enable choice of programmes that can contribute to broader aims, as well as achieving their specific objectives.

- the only answer required from the evaluation is about the average net effect of the intervention;
- programmes are genuinely simple and where one size really does fit all – the work involved in a realist evaluation is not warranted in such circumstances;
- the human and financial resources required to undertake a realist evaluation are not available.

In what circumstances is a realist approach more appropriate than other methods?

Experimental or quasi-experimental designs are often advocated for impact evaluation. However, they may not be feasible or desirable when:

- the programme has not been adequately developed and trialled (even in drug trials, RCTs are not used until there is a clear theory about what is expected to work, the drug itself has been developed, it is known whether the drug has negative effects and the effective dose is known);
- the scale of the intervention is too small for an RCT to be feasible (for example, during development of a new intervention);
- the programme is universal (or will reach 100% of its target group) and therefore no comparison group is available;
- there are ethical concerns about withholding an intervention from some people;
- the cost of undertaking an RCT (or similar design) is prohibitive.

A realist evaluation can be more appropriate than experimental or quasi-experimental designs in such situations. Data are analysed in realist impact evaluation through intra-programme, inter-group comparisons. Rather than comparing changes for participants who have undertaken a programme with a group of people who have not, as is done in random control or quasi-experimental designs, a realist evaluation compares whether a programme works differently in different localities (and if so, how and why) or for different population groups (for example, men and women, or groups with differing socio-economic status). Realist evaluations can be undertaken with small or large groups and with qualitative and/or quantitative data.

What circumstances make it easier to undertake a realist impact evaluation?

While realist ideas can be applied in most impact evaluation designs, some circumstances make it easier and others make it more difficult. Table 2 (p.9) lists the circumstances in which realist designs are easier, in the left hand column; the reason why the circumstance matters, in the centre column; and the strategy to address the difficulty should the circumstance not exist, in the right hand column.

Table 2. Circumstances and realist impact evaluation

Helpful circumstance	Implication of the circumstance for evaluation	Alternate strategies when circumstance is absent
Prospective or concurrent evaluation	CMO hypotheses can be developed before implementation, with the impact evaluation designed to test those hypotheses	For retrospective evaluations, stakeholders can be interviewed about context and mechanism, but analyses of outcomes for sub-groups will only be possible if data on sub-groups was collected
Implementation occurs across multiple sites or with multiple groups	Allows testing of 'for whom' / in what contexts	For small group / single site pilot programmes, in-depth analysis using qualitative data can be used to build hypotheses for testing in later, larger stages of the programme
Longer term projects with concurrent evaluation	Allows iteration and refinement of the evaluation design over time	Delimit the evaluation questions more tightly
Policy and programme personnel understand the role and value of programme theory	Time and resources are more likely to be allocated to develop and test programme theory.	Provide practical examples of the use of programme theory for the commissioner, policy personnel and programme personnel's own purposes and contexts
Data systems allow disaggregation of outcomes by sub-groups	Allows testing of programme theories	Post-hoc interviews with practitioners or subjects to identify differential outcomes for different sub-groups
'Objective' and quantifiable outcome indicators exist	Facilitates sub-group analysis of outcomes	Analyse qualitative outcomes data by sub-groups, according to theory
Mixed methods evaluation	Allows for quantitative indicators for outcomes, qualitative exploration of mechanisms.	For some designs, purely quantitative or purely qualitative data can be used
Clear initial programme theory	Programme theory is a pre-requisite for realist analysis	Work with programme designers and implementers to construct programme theory retrospectively
Evaluation team have existing skills in realist evaluation	Facilitates all aspects of the evaluation	One team member or external consultants provide mentoring in realist methodology
Adequate resources relative to programme size and complexity	Enables more comprehensive evaluation	Restrict the scope of the evaluation to a more limited range of questions

Designing and conducting a realist impact evaluation

This section describes the requirements and processes for designing and conducting a realist impact evaluation. These are similar to designing and conducting many other kinds of impact evaluations, so the focus here is on specific changes needed because of the nature of a realist approach.

Evaluation design is an iterative, rather than a linear process – it will be necessary to move back and forth between the stages until the best possible design has been achieved.

Establishing working relationships

Realist evaluations are intended to inform policy or practice, and collaboration with commissioners and stakeholders is likely to improve how useful the evaluation is. Working out the extent to which commissioners and programme staff can participate in the design process and whether and how they will contribute to undertaking the evaluation itself is useful as part of the design process.

Purpose

Clarifying the purposes for which the impact evaluation will be used provides the focus for the evaluation. The issue here is not 'what are the evaluation questions?', but 'for what will the answers be used?'

This means understanding the policy or practice issues that commissioners need to address. Is the aim to refine this programme in this setting – to increase its effectiveness, expand the range of target groups for which it is effective, or inform the choice of the range of programmes required in a policy domain? Is it to enable the programme to be adapted for scaling out to other settings? Is it perhaps to work out whether the programme approach can be applied to other policy questions?

Understanding the purposes will help to refine and prioritise the evaluation questions (see 6.3 below).

Overarching & subsidiary questions

The guiding questions for a realist evaluation need to reflect the principles of realist evaluation as well as the purpose for the evaluation. Some questions will be descriptive (saying what has happened and to what extent) and others will be explanatory (how or why something happened). Questions may include:

- For whom does the intervention work and not work, and why? (Note: this question does not ask 'who are the intended beneficiaries for this programme?'. It asks: *within* the intended beneficiaries, for which sub-groups is it more or less effective?)
  - Which sub-groups were reached by the programme?
  - How many people from which sub-groups actively participated?
  - What influenced whether sub-groups participated?
  - What were the outcomes for the various sub-groups?
  - What outcomes were expected and what were unexpected?
- In what respects does it work and not work for different groups?
  - For those who participated or were affected, do the types of outcomes achieved vary across sub-groups? Why?
- To what extent does it work or not work, for different groups or in different contexts?
  - How strong are the impacts for different sub-groups or in different contexts?
- When it works, how (i.e. by what mechanisms) does it work?
  - Did the expected mechanisms operate? For whom?
  - Were there unexpected mechanisms?
- When it doesn't work, why doesn't it work?
  - For whom did expected mechanisms not operate?
  - What features of context prevented anticipated mechanisms from firing?
- What matters about how it is done, in order for it to work?
  - What were the critical aspects of implementation, programme staffing, or organisational context that influenced how the programme operated?
  - Which aspects influenced whether, or which, mechanisms operated?
- What matters about the contexts into which it is introduced, in order for it to work?
  - What were the critical features of culture, belief systems, population group, history and so on that influenced whether or which mechanisms operated?

As with any evaluation, the questions will need to be prioritised: no evaluation will be able to answer every question that could be asked. The priority afforded to questions will need to reflect the purposes for which the evaluation will be used, but may also be influenced by the extent of existing knowledge about a programme or programme type.

Understanding the programme theory

Realist evaluation is explicitly theory-oriented: the purpose of a realist evaluation is as much to test and refine the programme theory as it is to determine the programme outcomes in a particular setting.

There are two distinct tasks in developing a realist programme theory. The first is to understand 'the basic programme theory'. The most basic format for programme theories is "If we do 'x', 'y' will happen, because…". Note that the 'because…' element of this structure is critical. Without it, there may be a theory of action ("We expect this to be done, or that this was done"), but there will not be a theory of change.

Realist evaluation assumes a primary type of programme theory – the 'context-mechanism-outcome' (CMO) hypothesis. The second task is to develop these hypotheses. The basic programme theory is refined by hypothesising answers to the four most basic realist questions:

1. *For whom* will this basic programme theory work and not work, and why?
2. In what *contexts* will this programme theory work and not work, and why?
3. What are the main *mechanisms* by which we expect this programme theory to work?
4. If this programme theory works, what *outcomes* will we see?

Because an intervention can trigger multiple mechanisms and the operation of mechanisms is context dependent, linked sets of hypotheses ("in this context, these mechanisms leading to 'x' outcomes; and in that context, those mechanisms leading to 'y' outcomes") are most likely to be generated. Generating hypotheses may require a workshop involving evaluators, commissioners and programme and policy staff. It may involve reading previous evaluations of related programmes to identify for whom, where and how they appeared to work. It may involve extrapolation from formal theories in the particular domain (for example, education theories in education, economic theories in economic development).

The results of hypothesis generation can usefully be constructed as a chart listing multiple C-M-O patterns (often referred to in realist literature as CMOC – a CMO configuration). Note, however, that these tables are not "lists of contexts" and "lists of mechanisms". Each CMO must be able to be read "as a sentence" across the rows of the chart: "In this context, that mechanism generates this outcome". Returning to our earlier free tuition example, "Poor parents who value education (C) are enabled (M) to send their children to school (O)".

The process of identifying and working through the implications of multiple hypotheses can feel overwhelming, and it is never possible to test all of them in one evaluation. CMOs need only be developed for the particular purposes and questions that have been identified in the earlier stages of the evaluation design. Only those which will be useful to the purpose of the evaluation should be tested.

It is preferable to develop CMO hypotheses prior to undertaking an evaluation because it enables data to be collected to test them. However in some circumstances, CMOs can be developed concurrently (for example, during the initial round of an evaluation, for testing during later stages of the evaluation) or retrospectively. In the latter case, the CMOs are a product of the evaluation, rather than a design tool.

Methods, instruments and data

As with any evaluation, the choice of methods and instruments to be used follows from the nature of data necessary to answer the evaluation questions. In a realist evaluation, the questions relate to the programme theory (for whom, how, in what contexts, and so on). Consequently the methods and instruments are selected to collect data that allows the theory to be tested and the questions to be answered.

Realist evaluation requires outcomes data: it is not possible to undertake a CMO analysis with no 'O(utcome)'. This requires adequate specification of the intended impacts of programmes (as does any form of impact evaluation) and appropriate financial resources, expertise, and staffing to collect that data.

A realist impact evaluation also requires that outcomes data can be disaggregated *according to sub-groups and contextual features identified in the realist programme theory.* This implies that:

- outcomes data are likely to be required at the unit (individual or organisation) level, and appropriate processes for protecting privacy and confidentiality will therefore be required;
- data must also be collected about relevant (or hypothesised to be relevant) features of context; and
- data about outcomes and features of context can be linked for analysis.

For example: if the realist programme theory suggests that a programme will work differently for men and women, then it will be important to analyse outcome data by sex. If it suggests it will work differently for urban and rural residents, then outcomes data will need to be able to be disaggregated by location. If, however, the critical distinction is actually between farm and non-farm workers, then occupation (rather than location) will be required. If the programme theory suggests that the programme will work better for women if it is delivered through health services, and for men if it is delivered through workplaces, then outcomes data will need to be able to be disaggregated both by sex and by type of delivery site (health services/work places).

As disaggregated analysis is easier with numerical data, some realists suggest that quantitative data be used for outcomes. However, realist analysis of qualitative data is also possible. This simply requires separating the data from the different sub-groups (e.g. men/women, urban/rural), analysing each set, and comparing the outcomes across the sets.

A realist impact evaluation needs more than data about contexts and outcomes. Understanding causation is a critical element of a realist evaluation, so data must be collected about hypothesized mechanisms as well.

Two strategies can be used to design data collection about mechanisms. The first strategy is to understand *why* the programme is expected to work differently for the various sub-groups or contexts identified in the programme theory. For example: why is it expected that the programme will work differently for men and women? Does it relate to physical characteristics such as height or strength? Economic characteristics such as typical occupations or control of household finances? Or cultural expectations of roles and cultural conditioning for those roles? The 'why' suggests the mechanisms and also suggests specific indicators that might be collected.

The second strategy is to use Pawson and Tilley's construct of programme mechanisms as involving "resources and reasoning". This means collecting information about (or related to) the 'reasoning' of whoever is expected to make a different decision as a result of the programme (e.g. beneficiaries, service providers). Data about reasoning can be collected quantitatively if there is a clear hypothesis about the aspect(s) of reasoning that are expected to come into play. For example, a mechanism may relate to the priority that parents give to children's education. An instrument may be developed which assesses the values that parents hold in relation to various aspects of education and the priority they afford to education relative to other priorities. The instrument might then be administered pre- and post-programme to ascertain whether values and priorities have changed as a result of the programme, or whether parents have simply been enabled to put their values into action. Where there are no clear hypotheses or where appropriate instruments do not exist, qualitative data can be collected to explore the reasoning of subjects in response to the resources provided through the intervention.

Analysis

Realist analysis uses data to test aspects of the programme theory. The usual form that analysis takes is intra-programme (all subjects are within the programme), inter-group comparisons, with the comparisons based on programme-theoretical constructs.

Imagine that a programme hypothesis is that the programme works by 'increasing social capital' (for example, building networks between farmers and markets, or strengthening relationships between peers who help each other in some way). During the development of programme theory, the type or nature of social capital (e.g. bonding capital, bridging capital) is identified. During evaluation design, appropriate indicators are selected (e.g. networks formed or strengthened, trust, or access to services). During analysis, the relationship between social capital before and after the programme, and programme outcomes, are analysed.

If the programme theory is correct, where there is no increase in the appropriate types of social capital, there should also be no change in outcomes. This might be the case, for example, where participants already had high levels of social capital, or where extreme isolation prevents new networks being established. Where social capital increased, however, there should also be a change in outcomes.

In the final stage of a realist evaluation, the findings are used to refine the programme theory. Refinements may include better understanding of what mechanisms are or how they actually work; identification of new mechanisms; better understanding of how factors in the context affect whether and which mechanisms operate; or a more refined understanding of the patterns of outcomes resulting from the interaction of context and mechanism.

Reporting

Standard criteria for the quality of an evaluation report apply to realist evaluations. However, a realist impact evaluation also requires that the realist nature of the investigation be made explicit. This implies:

- reporting the impact evaluation questions in realist format (as illustrated in the section on evaluation questions above);
- describing the initial programme theory and the initial CMO hypotheses for the programme;
- explaining what data was used to test which aspects of the programme theory;
- explaining the analytic techniques used with particular data sets;
- explicitly presenting and discussing disaggregated outcomes for different sub-groups identified in the programme theory;
- explicitly aligning evidence (on mechanisms, outcomes and context) against the programme theory, to provide a transparent basis for judgements about the programem (against the theory) and the theory (against the programme). *This is the primary distinguishing feature of a realist evaluation and its importance cannot be over-rated*. The aim is to assist the reader to understand HOW and WHY the pattern of impacts found by the evaluation exist;
- presenting the refined programme theory and its implications for policy and programmes. This section should link directly to the purposes for which the evaluation was commissioned.

Acknowledgements

Many thanks to Patricia Rogers and Irene Guijt for their review and to Jessica Sinclair Taylor and Steven Dickie for editing and design.

References

Office of Development Effectiveness (2012) Impact Evaluation: A Discussion Paper for AusAID Practitioners. Canberra: AusAID.

Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*. London: Sage.

UNIFEM (2009) Guidance Note on Carrying Out an Evaluability Assessment. New York: United Nations.

Wong, G., Westhorp, G., Pawson R, and Greenhalgh, T. (2012) *Realist Synthesis RAMESES Training Materials*. Available at: http://www.ramesesproject.org/media/Realist_reviews_training_materials.pdf