REALIST EVALUATION

OVERVIEW

Pawson and Tilley's (1997) starting point for setting out the realist approach to evaluation is to argue that the 'traditional' experimental evaluation is flawed because its attempt to reduce an intervention to a set of variables and control for difference using an intervention and control group strips out context. Instead evaluators need a method which 'seeks to understand what the program actually does to change behaviours and why not every situation is conducive to that particular process.' (Pawson & Tilley 1997: 11). They assume a different, 'realist' model of explanation in which 'causal outcomes follow from mechanisms acting in contexts' (Pawson & Tilley 1997: 58).

A mechanism explains precisely what it is about a programme that makes it successful. Mechanisms are not variables but accounts that encompass individual agency and social structures. They thus should 'reach down' to individual reasoning and 'reach up' to the collective resources embodied within a social programme that is being evaluated (Pawson and Tilley 1997). For Pawson and Tilley 'A mechanism is thus a theory – a theory which spells out the potential of human resources and reasoning.' (Pawson and Tilley 1997: 69). Astbury and Leeuw's (2013) definition of mechanisms as '...underlying entities, processes, or [social] structures which operate in particular contexts to generate outcomes of interest'

For Pawson and Tilley (1997) causal mechanisms and their effects are not fixed, but contingent on context. A programme will only 'work' if the contextual conditions into which it is inserted are conducive (Pawson and Tilley 1994). Programs are always introduced into a pre-existing social context and pre-existing structures enable or disable the intended mechanism of change (Pawson and Tilley 1997). This is to recognise the complexity of social interventions, using 'complexity' in its sociological sense to include the principle of non-linearity (small changes in inputs may, under some conditions but not others, produce large changes in outcome); the contribution of local adaptiveness and feedback loops; the phenomenon of emergence; the importance of path dependence; and the role of human agency (Marchal et al. 2012).

A substantial part of Pawson and Tilley's key texts (1994, 1997) in which they set out the case for scientific realist evaluation are given over to a discussion of causation. For Pawson and Tilley (1997) the model of causation adopted in 'traditional' experimental evaluation design is external, successionist causation. This is the idea that causation itself is unobservable but that it can be inferred from the basis of observation. Scientific realists prefer a model of generative causation that sees causation acting internally as well as externally. They do not therefore make predictions about the probability of an intervention leading to an outcome. This is because complex interventions are only semi-predictable (Lawson 1997, Marchal et al. 2012). Lawson's (1997) concept of demi-regularity is that human choice or agency is only semi-predictable because variations in patterns of behaviour are attributable partly to context. Human behaviour is not determined, but neither is it completely haphazard. There will be some patterning and therefore the best realist evaluation can offer is plausible explanations of what works for whom, in what circumstances and in what respects an intervention is more likely to succeed (Wong et al. 2013).

To be clear, Pawson and Tilley argue in favour of an experimental method. However, they reject the model of experiment based on a similar intervention and control group. They argue instead, following philosophers such as Bhaskar that the two essential elements of an experiment are triggering the mechanism being studied to make sure that it is active and preventing interference with the operation of the mechanism. In this model, rather than simply activating an independent variable and observing the outcome, the experimentalist's task is to manipulate the entire experimental system.

A particular concern has been about the ability of the approach to deal with complexity (Pederson and Rieper 2008, Blamey and Mackenzie 2007, Davis 2005). Critics argue that scientific realism was developed partly in the field of crime reduction where programmes under evaluation were relatively small scale, operating at a local level, targeted on distinct groups and involving relatively few stakeholders. However, Pederson and Rieper in their own work and through referencing that of others (e.g. Davis 2005) demonstrate how the scientific realist approach can be adapted to regional and national level policies and programmes that are more complex.

KEY ELEMENTS OF METHODOLOGY

There is still much debate about exactly how to undertake a realist evaluation. Marchal et al. (2012) in their review of realist evaluation in health systems research found significantly different approaches being used. These included different ways of treating mid-range theory, different understandings of 'mechanisms' and 'context' and different approaches to handling context-mechanism-outcome configurations. While, at the time of writing, a similar review did not exist for widening participation in higher education it is likely that there are also different approaches to realist evaluation being used within the sector. It is not therefore possible to set out a precise set of agreed steps that an evaluator should move through when undertaking a realist evaluation. Nevertheless it is possible to set out some key elements of a realist evaluation.

Mechanisms

Wong et al. (2013) suggest that one way to identify a programme mechanism is to reconstruct, in imagination, the reasoning of participants or stakeholders. They also note that:

- Mechanisms cannot be seen or measured directly (because they happen in people's heads, or at different levels of reality than the one being observed).
- There will potentially be a large number of mechanisms and the role of the realist researcher is to identify the 'main mechanisms', which they define as 'those that are common and significant enough to contribute to the pattern of outcomes of the intervention' (Wong et al. 2013: 6).
- The 'causes' of outcomes are not simple, linear or deterministic. This is partly because programmes often work through multiple mechanisms and partly because a mechanism is not inherent to the intervention, but is a function of the participants and the context. Consequently the same intervention can trigger different mechanisms for different participants, even within one location.
- Mechanisms are context-sensitive.

Context

Wong et al. (2013) note that context can take on a multitude of forms including:

- Broad social or geographical features (for example the country in which an intervention operates and its cultures)
- Features affecting the implementation of programs (for example whether the program occurs in a prison, a hospital or health service, whether there is adequate funding, the qualifications of staff).
- The make-up of the participants on a program or the different population profiles of locations in receipt of an intervention.
- Conditions in which subjects seek to enact their choices (graduates of a vocational training program will find it easier to get work in a context of high employment; recipients of a housing subsidy will find it harder to use that subsidy in a context of housing shortages.)

But, Wong et al. (2013: 9) are also clear that just because context can take a multitude of forms it is not necessary to list the 'infinite potential 'surrounds' to an intervention'. Instead:

> *'What matters is developing an understanding of how a particular context acts on a specific program mechanism to produce outcomes – how it modifies the effectiveness of an intervention.'*
>
> **(Wong et al. 2013: 9)**

Context-Mechanism-Outcome configurations

Pulling these elements together, the scientific realist evaluator always constructs their explanation around three vital ingredients of context, mechanism and outcome, which Pawson and Tilley refer to as context–mechanism–outcome configurations:

> *The basic task of social inquiry is to explain interesting, puzzling, socially significant regularities… Explanation takes the form of positioning some underlying mechanism… which generates the regularity and thus consists of propositions about how the interplay between structure and agency has constituted the regularity. Within realist investigation there is also investigation of how the workings of such mechanisms are contingent and conditional, and thus only fired in particular local, historical or institutional contexts…*
>
> **(Pawson and Tilley 1997: 71)**

For Pawson and Tilley (1994, 1997) traditional experiments misunderstand what makes programmes work: 'Programmes cannot be considered as some kind of external, impinging 'force' to which subjects 'respond'.' (Pawson and Tilley 1994: 294). Instead social programmes are social systems involving an interplay between individual and institution, or, in the language of Giddens (1984): agency and structure. Thus, it is not programmes which work but rather people who co-operate and choose to make them work. But, scientific realists do not adopt the same formulation as constructivists. Rather they see people's choices as constrained by social structures:

> *[P]rogrammes 'work', if subjects choose to make them work and are placed in the right conditions to enable them to do so. This process of 'constrained choice' is at the heart of social and individual change to which all programmes aspire…*
>
> **(Pawson and Tilley 1994: 294)**

Undertaking a scientific realist evaluation

The starting point is theory and 'empirical work in programme evaluation can only be as good as the theory which underpins it' (Pawson and Tilley 1997: 83). Thus, a scientific realist evaluation should not be data-driven, but theory-driven. To give a practical example: the subject of an interview would be the evaluator's theory, which the person being interviewed is asked to confirm or falsify and refine (Pawson and Tilley 1997: 155). The interview relationship has been described as a teacher-learner cycle in which the interviewee is taught the programme theory that is being tested and, having learned about the theory is able to teach the evaluator about the components of the programme (Pawson and Tilley 2004). For initial theory-gleaning interviews are likely to focus on practitioners.

Although both quantitative and qualitative data have a place and scientific realism is 'method neutral' (Marchal et al. 2013) there will generally be more emphasis on qualitative data that allows for theory to be explored and semi-structured interviews tend to be particularly common (Manzano 2016).

To be able to build explanations, data collection should be iterative (Manzano 2016). The ideal empirical evaluation would therefore collect 'before' and 'after' data to give an overall picture of outcomes, but thereafter, more attention would be given to gaining data which tapped mechanism and contextual variation (Pawson and Tilley 1994). Thus, there is a strong assumption that there will be multiple strands of data collection, with, for example, semi-structured interviews being supplemented with observations and/or analysis of quantitative data.

The standard scientific realist data matrix would make comparisons of variation in outcome patterns across groups, but those groups would not be experimental and control groups. Instead they would be defined by the mechanism/context framework with the evaluator running a systematic range of comparisons across a series of studies to understand which combination of context and mechanism works best. (Pawson and Tilley 1994)

MULTI-METHOD APPROACHES

Scientific realism can provide a framework within which some of the other 'small n' methodologies sit. Process Tracing, for instance, draws on many of the same underlying assumptions about causality, the nature of the social world (ontology) and the status of knowledge of the social world (epistemology).

Commentators have also noted that realist evaluation, as a form of theory-led evaluation, bears a close resemblance to the theories of change approach. However, Blamey and Mackenzie (2007: 452) argue that 'Theories of Change and Realistic Evaluation may both be from the same stable, [but] they are in practice very different horses'. For Blamey and Mackenzie, one of the key differences is that in a 'theories of change' approach 'theory' is articulated by a wide range of stakeholders, whereas in realist evaluation it is the evaluator who articulates the theory.

RESOURCES REQUIRED

Skill set for evaluators

In the scientific realist paradigm the evaluator is a researcher and theorist, with a detailed understanding of the programme being evaluated and able to construct mid-level theories (groups of context-mechanism-outcome configurations) for subsequent testing.

Although scientific realist evaluation can incorporate both quantitative and qualitative data collection, qualitative data collection, and in particular the semi-structured interviews tend to be most common (Manzano 2016). However, scientific realism involves a particular conception of the interview as a teacher-learner cycle Pawson and Tilley 2004, see above). The approach to interviewing required in a scientific realist interview may contradict the standard research methods training that an evaluator has received. Whereas the prevailing approach to interviewing is to maintain neutrality and avoid leading questions, a realist evaluation interview in which exploration of theory is the aim is likely to be led by the interviewer and aims at 'assisted sensemaking' (Manzano 2016). With a focus on the interview in realist evaluation, Manzano emphasises, not so much particular training or skills, but the idea of the researcher learning the 'craft' of interviewing, suggesting that this is an approach that takes time and experience to develop.

Resource implications

While there are no set rules for how much data should be collected, it is accepted that a realist evaluation should aim to collect large amounts of data (Manzano 2016: 348). Manzano explains that 'substantial amounts of primary or secondary data are needed – even when the sample is small – to move from constructions to explanation of causal mechanisms.' This is because the unit of analysis is not the person, but the events and processes around them and 'every unique programme participant uncovers a collection of micro events and processes, each of which can be explored in multiple ways to test theories' (ibid.). It is not possible to quantify what amounts to 'large amounts of data' in the abstract, but collecting multiple sources of data on each case suggests days, rather than hours of work.

CASE STUDY

Formby et al. (2020) describe the realist evaluation of Go Higher West Yorkshire (GHWY) Uni Connect – an initiative to reduce educational inequalities through collaborative widening participation (WP) outreach. It contributes to wider debates on widening participation policy through demonstrating how (HEPOs) normalised 'progression' based on community and learners' needs. The evaluation aim was to understand the differences in approach of Higher Education Progression Officers (HEPOs) working in community settings to normalise 'progression' based on community and learners' needs.

Three programme theories were developed around the ideal practice of HEPO staff, based on a series of initial interviews and focus groups conducted with GHWY staff involved in the Uni Connect programme. These were:

- Good quality continuing professional development (CPD) will equip school/college-based staff with the skills and information to support young people to make informed choices.
- Dedicated progression staff in schools/colleges will have more time to invest in young people and support them in planning for their future.
- Facilitating the delivery of outreach activity aimed at helping young people to make informed choices.

The programme theories were then split into several hypothetical Context – Mechanism – Outcome configurations through analysis of programme documentation, literature reviews on effective WP outreach and discussion with stakeholders. The CMOs suggested the need for individual responses from HEPO staff and management to uncover wider cultural models relating to the normalisation of WP in school/college institutions, as well as wider societal/community factors outside the institution that may impact the HEPO. The evaluators undertook realist interviews with managers of HEPOs and focus groups with HEPO front-line staff to determine and refine valid CMO configurations. Both data-collection methods were chosen because they are effective at identifying contexts and mechanisms that produce variant outcomes.

The evaluation showed that HEPOs both complemented existing arrangements in settings that already practised WP and introduced new WP activity that shifted the wider cultural practice in settings where WP resources had been introduced for the first time.

Reference

Formby, A., Woodhouse, A. and Roe, F. (2020) 'A Presence in the Community': Developing Innovative Practice through Realist Evaluation of Widening Participation in West Yorkshire', *Widening Participation and Lifelong Learning* Vol.22(3) pp. 173-186 DOI:10.5456/WPLL.22.3.173

RESOURCES

Web resources

An interesting interview about realist evaluation with Ray Pawson is available on the Vimeo platform. The interview is divided into four parts and the first one is available here: https://vimeo.com/84215487

The RAMESES II project, funded by the NIHR developed quality and reporting standards and resources and training materials for realist evaluation. These are available on-line at: https://www.ramesesproject.org/Home_Page.php

There is a Supplementary Guide on Realist Evaluation issued as part of the Magenta Book 2020, available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879435/Magenta_Book_supplementary_guide._Realist_Evaluation.pdf

Key reading

Pawson and Tilley's influential and widely cited book on scientific realist evaluation is a good starting point for exploring scientific realism:

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

Ana Manzano explores the craft of realist interviewing and, in doing so, helps to illuminate key elements of the conduct of a realist evaluation:

Manzano A. (2016) The craft of interviewing in realist evaluation. *Evaluation*. Vol.22(3) pp.342-360. Doi:10.1177/1356389016638615

Pederson and Rieper explore a key criticism of realist evaluation, namely its ability to deal with complexity:

Pederson LH and Rieper O (2008) 'Is Realist Evaluation a Realistic Approach for Complex Reforms?', *Evaluation*, 14(3) pp.271-93.

Realist evaluation is closely related to theories of change and Blamey and Mackenzie's article in exploring the similarities and differences also highlights important elements of the realist approach to evaluation:

Blamey A and Mackenzie M (2007) 'Theories of Change and Realistic Evaluation: Peas in a Pod or Apples and Oranges?' *Evaluation* 13 pp. 439-455.

Further references

Blamey A and Mackenzie M (2007) 'Theories of Change and Realistic Evaluation: Peas in a Pod or Apples and Oranges?' *Evaluation* 13 pp. 439-455.

Davis P (2005) 'The Limits of Realist Evaluation: Surfacing and Exploring Assumptions in Assessing the Best Value Performance Regime', *Evaluation* 11 pp.275–95.

Giddens A (1984) *The Constitution of Society*, Cambridge: Polity Press.

Lawson T. (1997). *Economics and Reality*. Routledge: London.

Manzano A. (2016) The craft of interviewing in realist evaluation. *Evaluation*. Vol.22(3) pp.342-360. Doi:10.1177/1356389016638615

Marchal B, Belle S, Olmen J, Hoerée T and Kegels G (2012) 'Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research', *Evaluation* 18(2) pp. 192-212.

Pawson, R, Tilley, N (2004) *Realistic evaluation. British Cabinet Office*. Available at: http://www.communitymatters.com.au/RE_chapter.pdf [accessed 6th September 2021]

Pawson R and Tilley N (1994) 'What works in evaluation research?' *British Journal of Criminology* 34 pp.291-306.

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

Pederson LH and Rieper O (2008) 'Is Realist Evaluation a Realistic Approach for Complex Reforms?', *Evaluation*, 14(3) pp.271-93.

Wong, G., Westhorp, G., Pawson, R. and Greenhalgh, T (2013) Realist Synthesis RAMESES Training Materials.