RESEARCH · MARKET AND BUSINESS RESEARCH

Realist Evaluation Approaches for Context-Mechanism-Outcome Knowledge

How realist evaluation asks what works, for whom, in what circumstances and why: the context-mechanism-outcome configuration, the realist evaluation cycle, reporting standards, current debates and applications to organisational initiatives.

ModQu research article · 1,678 words · 13 references (APA 7th edition)

Many evaluations ask whether a program works. Realist evaluation asks a more detailed question: what works, for whom, in what circumstances, and why? The approach was set out by Pawson and Tilley (1997) in *Realistic Evaluation*, and it has since been widely used in health services, criminal justice, education, international development and management research. Its central tool is the context-mechanism-outcome configuration, a way of stating how a program produces its effects and under which conditions. This article explains the ideas behind realist evaluation, how a study is carried out, the debates about its key concepts, and how the same thinking can be applied to business and organisational initiatives.

Why ask "for whom and in what circumstances"?

Programs rarely work in the same way everywhere. A mentoring scheme that helps new employees in one office may make no difference in another. A crime prevention measure that reduces theft in one car park may have no effect a few kilometres away. Conventional outcome evaluations often report an average effect across all sites and participants. An average can hide a program that works very well for some people and not at all, or even harmfully, for others.

Pawson and Tilley (1997) argued that this is not a technical inconvenience but a basic feature of social programs. Programs do not act on people directly. They offer resources, such as information, money, advice, opportunities or constraints, and people respond to those resources in different ways depending on their circumstances. It is people's reasoning and responses, not the program itself, that produce outcomes. To understand whether a program will work somewhere new, an evaluator therefore needs to know how it works and what conditions it depends on.

This matters most when a decision-maker wants to transfer a program. A positive average effect in one place does not show that the program will work somewhere else, because the new setting may lack the conditions that made it work. Pawson (2006) argued that what transfers between settings is usually not the program itself but the explanation of how it works. A policy-maker who knows which mechanism produced the result, and which conditions it needed, can judge whether those conditions exist in the new setting, or whether the program needs to be adapted before it is introduced.

The realist view of causation

Realist evaluation draws on scientific realism, a position in the philosophy of science associated with Bhaskar (1975). Realism holds that the world contains real causal mechanisms that are often not directly observable, and that whether a mechanism produces an effect depends on the conditions in which it operates. Gunpowder has the power to explode, but it explodes only when it is dry, compacted and ignited. The causal power is real; whether it is triggered depends on the context.

Pawson and Tilley (1997) contrasted this **generative** view of causation with what they called the **successionist** view, in which causation is inferred from regular associations between a cause and an effect, as in a comparison between a treatment group and a control group. They argued that experimental designs answer whether an average effect occurred, but strip away the context that explains why. Pawson (2013) developed this critique further, while also accepting that realist evaluations can use experimental and quantitative data as part of a wider explanatory approach.

Context, mechanism and outcome

Realist evaluation expresses its explanations as context-mechanism-outcome configurations, usually written as C + M = O: in this context, this mechanism is triggered and produces this outcome.

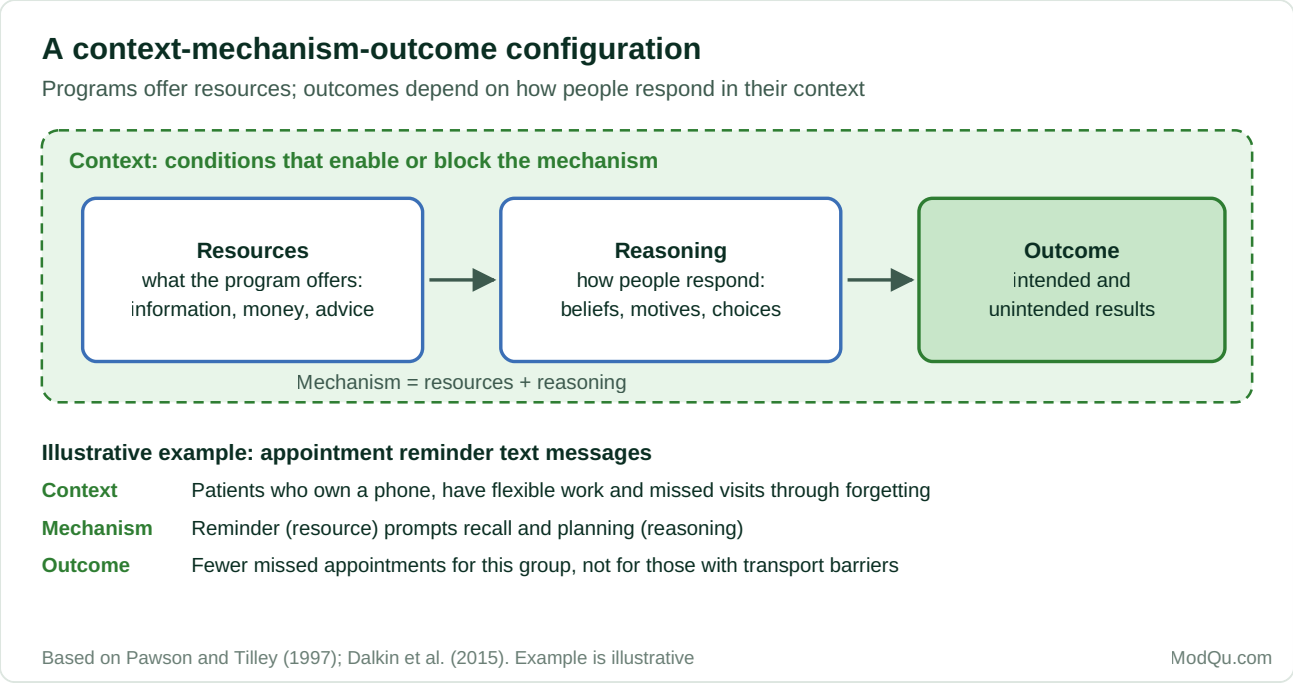


Figure 1. A context-mechanism-outcome configuration, with the mechanism divided into the resources a program offers and the reasoning of the people who receive them (Dalkin et al., 2015). The worked example is illustrative. Diagram by ModQu.

Context refers to the conditions that affect whether a mechanism is triggered. It can include characteristics of participants, staff, organisations, relationships, cultures, rules and wider social or economic conditions. Context is not simply the location of a program. Pawson and Tilley (1997) stress that the relevant context is whatever influences the operation of the mechanism.

Mechanism is the hardest concept to pin down. Pawson and Tilley (1997) describe mechanisms as what it is about a program that makes it work, generally in terms of how participants respond to the resources it provides. Dalkin et al. (2015) proposed making this explicit by dividing a mechanism into two parts: the **resources** introduced by the program and the change in participants' **reasoning** in response to them. They also suggested that mechanisms are better thought of as operating to varying degrees, rather than simply being switched on or off.

Outcome refers to the intended and unintended results of the mechanism, which may be short-term changes in behaviour or longer-term effects.

Pawson and Tilley's (1997) best-known example concerned closed-circuit television (CCTV) in car parks. They noted that CCTV could reduce car crime through several different mechanisms. It might deter offenders who fear being caught; it might lead to more legitimate users of the car park, whose presence increases natural surveillance; or it might help security staff respond more quickly. Which mechanism operates, and whether any does, depends on context: the layout of the car park, whether cameras are monitored, the times when crimes occur, and the kind of offenders involved. An evaluation that only asked whether car crime fell would not distinguish between these explanations, and would offer little guidance to someone deciding whether to install CCTV elsewhere.

How a realist evaluation is carried out

Pawson and Tilley (1997) describe the realist evaluation cycle in four stages, shown in Figure 2.

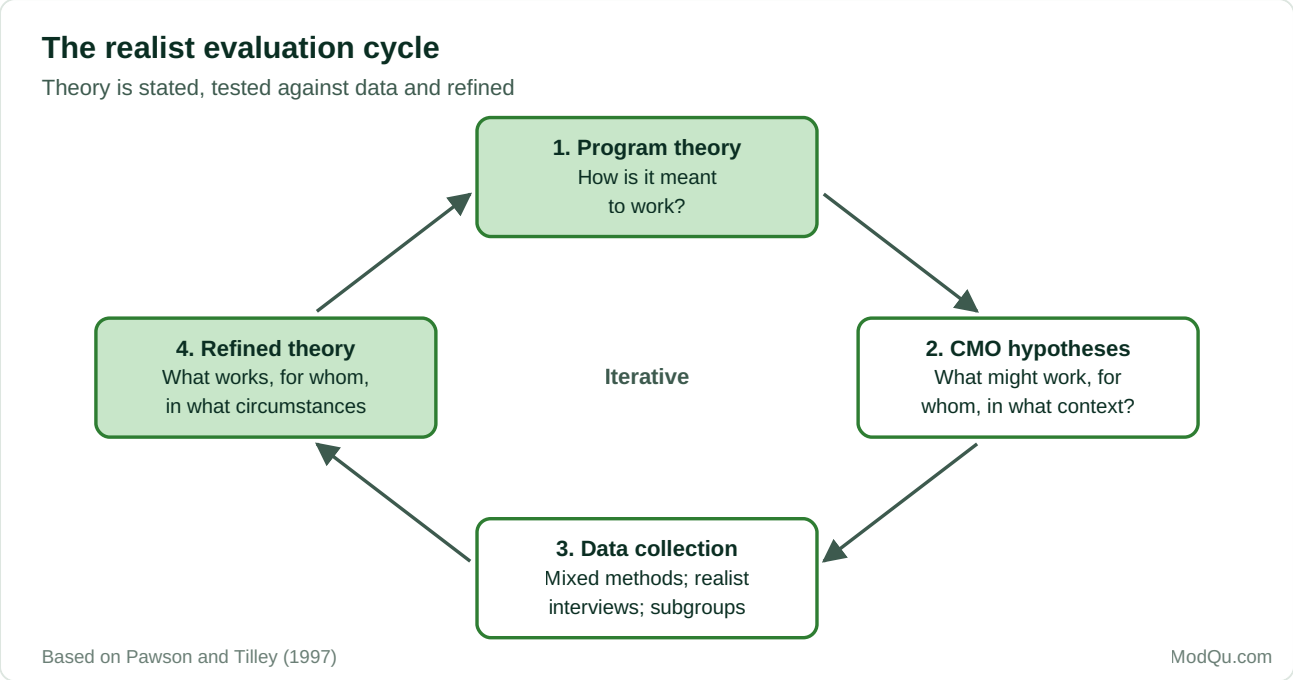


Figure 2. The realist evaluation cycle. Each study refines the program theory, which then informs the next round of evaluation. Based on Pawson and Tilley (1997); diagram by ModQu.

1. Develop an initial program theory. The evaluator sets out how the program is expected to work, drawing on program documents, the views of designers and staff, and existing research. Pawson et al. (2005) developed realist review, a form of evidence synthesis that can be used at this stage to assemble theories from previous studies.

2. Formulate context-mechanism-outcome hypotheses. The program theory is broken down into testable configurations: for whom, under what conditions and through what reasoning the program is expected to produce each outcome.

3. Collect and analyse data. Realist evaluation does not have a fixed method. It typically combines quantitative data on outcomes, broken down by subgroup and site, with qualitative data on how people responded to the program. Manzano (2016) describes the realist interview, in which the interviewer presents the program theory to participants and invites them to confirm, refute or refine it, rather than simply asking open questions about their experiences.

4. Refine the program theory. The findings are used to revise the configurations, specifying more precisely what works for whom and in what circumstances. The refined theory becomes the starting point for the next evaluation.

A realist evaluation of a large health service reform in London illustrates the approach. Greenhalgh et al. (2009) studied a program designed to modernise stroke, kidney and sexual health services. Rather than giving a single verdict on the program, they identified the mechanisms by which change occurred, or failed to occur, in different services, and how these depended on contextual conditions such as existing relationships and organisational history.

Reporting standards

As the approach spread, concerns grew about inconsistent practice. Marchal et al. (2012) reviewed published realist evaluations in health systems research and found that many applied the approach unevenly, with particular difficulty in defining mechanisms and distinguishing them from context. In response, the RAMESES projects developed quality and publication standards: first for realist syntheses (Wong et al., 2013) and then for realist evaluations (Wong et al., 2016). The RAMESES II standards set out 20 items to report, including the initial

program theory, the rationale for using a realist approach, the data collection methods and how the configurations were developed and refined.

Debates about the core concepts

The key concepts of realist evaluation remain the subject of debate. Lacouture et al. (2015) reviewed how the concept of mechanism was used in published realist studies and found a wide range of definitions, which made studies hard to compare. Greenhalgh and Manzano (2022) reviewed how context was understood in realist evaluations and syntheses. They found two broad framings: context as a set of observable features of a setting, and context as something relational and dynamic that shapes, and is shaped by, the program over time. The distinction matters in practice: the first treats context as a list of conditions to record, while the second requires following how a setting changes as the program is delivered.

Practical criticisms also recur. Realist evaluations can be time-consuming, because they require theory development, mixed methods and iterative analysis. Separating context from mechanism in real data is often difficult. And because realist evaluation produces conditional explanations rather than a single effect size, its findings can be harder to summarise for decision-makers who want a clear yes or no.

Applying realist thinking in organisations

Realist evaluation is most often used for public programs, but its questions are directly relevant to business initiatives. A company rolling out a new performance management system, a sales incentive scheme or a customer onboarding process

across many branches faces the same issue: the initiative will work better in some places than others, and knowing why matters more than knowing the average effect.

Useful realist practices for organisational evaluation include:

- **Write the program theory down.** State how the initiative is supposed to change behaviour. For example, a sales dashboard is expected to work because managers can see which accounts need attention and will act on that information.
- **Name the conditions it depends on.** The dashboard mechanism depends on managers having time to act, trusting the data and being rewarded for follow-up.
- **Break results down.** Compare outcomes across sites, teams and customer groups rather than reporting only the overall average.
- **Ask people about their reasoning.** Interviews with staff in high- and low-performing sites can reveal whether the intended mechanism is operating, and what prevents it where it is not.
- **Revise before scaling.** Use the refined theory to decide where and how to extend the initiative.

The result is evidence that helps decision-makers predict where an initiative is likely to work and what needs to be in place for it to do so. That is the kind of knowledge realist evaluation was designed to produce: not whether a program works in general, but how, for whom and under what conditions (Pawson, 2006).

References

Bhaskar, R. (1975). *A realist theory of science*. Leeds Books.

Dalkin, S. M., Greenhalgh, J., Jones, D., Cunningham, B., & Lhussier, M. (2015). What's in a mechanism? Development of a key concept in realist evaluation. *Implementation Science, 10*, Article 49. https://doi.org/10.1186/s13012-015-0237-x

Greenhalgh, J., & Manzano, A. (2022). Understanding 'context' in realist evaluation and synthesis. *International Journal of Social Research Methodology, 25*(5), 583–595. https://doi.org/10.1080/13645579.2021.1918484

Greenhalgh, T., Humphrey, C., Hughes, J., Macfarlane, F., Butler, C., & Pawson, R. (2009). How do you modernize a health service? A realist evaluation of whole-scale transformation in London. *The Milbank Quarterly, 87*(2), 391–416. https://doi.org/10.1111/j.1468-0009.2009.00562.x

Lacouture, A., Breton, E., Guichard, A., & Ridde, V. (2015). The concept of mechanism from a realist approach: A scoping review to facilitate its operationalization in public health program evaluation. *Implementation Science, 10*, Article 153. https://doi.org/10.1186/s13012-015-0345-7

Manzano, A. (2016). The craft of interviewing in realist evaluation. *Evaluation, 22*(3), 342–360. https://doi.org/10.1177/1356389016638615

Marchal, B., van Belle, S., van Olmen, J., Hoerée, T., & 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), 192–212. https://doi.org/10.1177/1356389012442444

Pawson, R. (2006). *Evidence-based policy: A realist perspective*. Sage.

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

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review: A new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy, 10*(Suppl. 1), 21–34. https://doi.org/10.1258/1355819054308530

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage.

Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J., & Pawson, R. (2013). RAMESES publication standards: Realist syntheses. *BMC Medicine, 11*, Article 21. https://doi.org/10.1186/1741-7015-11-21

Wong, G., Westhorp, G., Manzano, A., Greenhalgh, J., Jagosh, J., & Greenhalgh, T. (2016). RAMESES II reporting standards for realist evaluations. *BMC Medicine, 14*, Article 96. https://doi.org/10.1186/s12916-016-0643-1