**INTRODUCTION TO REALISTIC EVALUATIONS**

**by Simon Moss**

|  |
| --- |
| **Introduction** |

Researchers often conduct studies to examine the effect of some treatment, or intervention, on the health or wellbeing of participants. Often, researchers utilize a design called a randomized control trial, or experiment, in which

* half the participants are randomly assigned to the condition in which they receive the treatment or intervention, such as some medication
* the other participants are randomly assigned to the control condition in which they do not receive the treatment or intervention; they might receive a placebo instead, for example
* researchers measure the health or wellbeing of participants, often before and after the treatment and intervention
* if the participants who are assigned the treatment demonstrate greater improvements in health or wellbeing than do the participants who are assigned to the control condition, the intervention is assumed to be effective.

**Limitations of the randomized control trial for complex social interventions**

However, in some circumstances, randomized control trials are not suitable. Specifically, when the intervention is complex and social, randomized control trials are inappropriate. That is

* interventions are complex when they entail a blend of distinct features—such as a series of activities—rather than one feature—such as a single drug
* interventions are social when they depend on the choices and decisions of humans, including both the participants and the researchers.

The following table illustrates some examples of complex, social interventions.

|  |  |
| --- | --- |
| Example | Details |
| Community mental health centre to help people recover from mental illness | * The center provided useful information about recovery from mental health * Individuals who were recovering from mental illnesses were granted opportunities to support each other * Individuals who were recovering from mental illnesses could participate in structured activities * This support was also extended to the carers of these individuals |
| A program in which parents massage their children | * Parents were encouraged to massage their infant children to improve the bond * This procedure seems to be effective in only a subset of relationships * This approach was effective in parents who exhibited moderate risk of problems, but only if the intervention comprised many of the features or mechanisms that are shown to be effective (see Underdown, Norwood, & Barlow, 2013) |

In these circumstances, for several reasons, randomized control trials are not useful—and neither are meta-analyses. The following table outlines these reasons

|  |  |
| --- | --- |
| Limitations of randomized control trials when interventions are complex and social | Details |
| Cannot separate treatment and control conditions | * The intervention is usually embedded within a social environment, such as an organization or community * In these environments, researchers cannot as readily determine who receives the treatment |
| The features of these interventions vary | * Researchers cannot readily control which features of the intervention participants receive * These features may also evolve over time |
| Participants vary in how they respond to the features of this intervention | * Some participants, for example, might disregard some features of the intervention |

**Aim of researchers when the intervention is complex and social**

Partly because of these complications, researchers do not merely want to ascertain whether some intervention is effective. Instead, when the intervention is complex and social, their purpose is to understand

* which features of the intervention are beneficial and detrimental
* what are the beneficial and detrimental effects of this intervention
* in which circumstances, and for which people, is the intervention most beneficial or detrimental
* why is the intervention beneficial or detrimental in these circumstances.

Ray Pawson and Nick Tilley developed an approach, called realistic evaluations, to answer these questions (e.g., Pawson & Tilley, 1997). This document outlines this approach.

|  |
| --- |
| **Overview of realistic evaluations** |

This section briefly delineates the key features of realistic evaluations. Realistic evaluations comprise four phases.

* First, using previous literature and experience, the researchers construct hypotheses about the circumstances in which the intervention might be most effective and the reasons this intervention might be most effective
* Second, the researchers collect qualitative data, such as interviews, quantitative data, such as surveys, or both to assess these hypotheses
* Third, the researchers analyze the data to assess these hypotheses
* Finally, the researcher interprets the data—to ascertain which features of the intervention are beneficial and detrimental and in what circumstances as well as why
* The researchers may then utilize these insights to update the hypotheses and initiate this sequence of procedures again.

**Constructing hypotheses: CMO configurations.**

During the first phase, to construct hypotheses, researchers utilize a scheme called context, mechanism, outcome—or CMO—configurations. The following table defines these three terms.

|  |
| --- |
| Definition |
| **Context**   * The characteristics that vary across participants or circumstances that might affect the impact of this intervention. * These characteristics or circumstances initiate the changes or mechanisms that culminate in some outcome or effect |
| **Mechanisms:**   * The sequence of changes that explain how some intervention generates some outcome. * Mechanisms can include a sequence of activities, resources, as well as the behaviors or responses of humans. |
| **Outcome**   * The effects of some intervention |

In essence, researchers strive to characterize the mechanisms—the theories or accounts—that explain how the intervention generates some outcome or benefit. In addition, researchers assume the context, such as the characteristics of participants or settings, determine which mechanisms or changes are activated. To illustrate, suppose researchers wanted to ascertain whether various opportunities to work in industry enhance the motivation of PhD and Masters by Research candidates. The following table presents some contexts, mechanisms, and outcomes that correspond to this example. As this table shows

* Contexts may refer to a subset of participants or settings [C1 to C4]—or a characteristic that varies across participants or settings [C5]. Contexts can also refer to various kinds of interventions [C6].
* Mechanisms may refer to explanations of why some intervention could be effective [M1 to M4] as well as the key features of these interventions [M5 to M6] or responses of participants
* For educational interventions, Kirkpatrick and Kirkpatrick (1994) differentiate four kinds of outcomes: the satisfaction of participants, the knowledge, skills, or attributes that participants have acquired, changes in the behavior or performance of participants, and the final outcomes—such as profitability

|  |  |  |
| --- | --- | --- |
| Contexts | Mechanisms | Outcomes |
| * PhD and Masters by Research candidates who will finish within a year but have not worked extensively in industry [C1] * PhD and Masters by Research candidates who will finish within a year but have worked extensively in an industry that diverges from their research [C2] * PhD and Masters by Research candidates who will finish within a year and have worked extensively in the industry in which they conduct research [C3] * PhD and Masters by Research candidates who will not finish within a year [C4] * The degree to which candidates live close to their university [C5] * Whether the industry is a government department or private firm [C6] | * Helping candidates envisage their career more vividly [M1] * Helping candidates learn the language of this industry [M2] * Attracting feedback about which of their skills are beneficial to industry [M3] * Attracting feedback about their shortfalls [M4] * An industry supervisor that encourages candidates to reflect upon their goals [M5] * An industry supervisor that is experienced in academia [M6] | * Motivation to complete the thesis [O1] * Confidence about future careers [O2] |

To understand this CMO model, scholars often refer to the analogy of a man, standing on a tennis court, with a tennis ball in their hand. When he opens his hand, the tennis ball drops. This simply analogy underscores some vital insights

* the opening of this hand does not cause the tennis ball to drop: after all, the tennis ball would not drop in space or in the ocean
* on the tennis court, the drop can be ascribed to gravity
* in an ocean, the tennis ball might rise because of buoyancy
* several mechanisms—such as gravity and buoyancy—operate on the tennis ball simultaneously
* several contexts—such as space or the ocean—affect which mechanisms prevail
* the opening of this hand is comparable to an intervention and differs from the mechanisms

Initially, to identify these contexts, mechanisms, and outcomes, the researchers might

* consult the theoretical and empirical literature on this topic
* observe this intervention or similar interventions
* interview relevant stakeholders, such as the managers who want to implement this intervention.

As the research unfolds, the researchers may then refine and update these contexts, mechanisms, and outcomes.

**Collecting data**

The researcher then attempts to collect data to assess

* which mechanisms are related to the various outcomes
* how the contexts affect which mechanisms are related to the various outcomes

Researchers might utilize many sources of data, such as observations, interview, surveys, and other measures. To illustrate, the following table specifies how researchers could measure the contexts, mechanisms, and outcomes that were stipulated in the previous table.

|  |  |  |
| --- | --- | --- |
| Contexts, Mechanisms,  and Outcomes | Source of data | Example |
| **Context** |  |  |
| PhD and Masters by Research candidates who will finish within a year but have not worked extensively in industry [C1] | * Official candidate records * CV |  |
| PhD and Masters by Research candidates who will finish within a year but have worked extensively in an industry that diverges from their research [C2] | * Official candidate records * CV |  |
| PhD and Masters by Research candidates who will finish within a year and have worked extensively in the industry in which they conduct research [C3] | * Official candidate records * CV |  |
| PhD and Masters by Research candidates who will not finish within a year [C4] | * Official candidate records |  |
| The degree to which candidates live close to their university [C5] | * Official candidate records |  |
| Whether the industry is a government department or private firm [C6] | * Interview of supervisors |  |
| **Mechanisms** |  |  |
| Helping candidates envisage their career more vividly [M1] | * Interview of candidates | * “What do you imagine when you consider your future career? How vivid is this image?” |
| Helping candidates learn the language of this industry [M2] | * Interview of candidates | * “What terms did you learn during your industry experience?” |
| Attracting feedback about which of their skills are beneficial to industry [M3] | * Interview of candidates | * “Which of your skills did you discover might be more helpful to your future career?” |
| Attracting feedback about their shortfalls [M4] | * Interview of candidates | * “Which of your skills did you discover you might need to develop to attract the roles you like?” |
| An industry supervisor that encourages candidates to reflect upon their goals [M5] | * Interview of supervisors | * “How did you provide feedback to candidates?” |
| An industry supervisor that is experienced in academia [M6] | * Interview of supervisors |  |
| **Outcomes** |  |  |
| Motivation to complete the thesis [O1] | * Survey of candidates before and after the intervention | * On a scale of 1 to 10, how engaged do you feel now while working on your thesis |
| Confidence about future careers [O2] | * Survey of candidates before and after the intervention | * On a scale of 1 to 10, how confident are you that you will be able secure the job you like |

**Analyzing and interpreting the data**

After collecting the data, researchers need to conduct analyses to assess the context-mechanism-outcome configuration, developed previously. The quantitative data can be subjected to various statistical tests. For example

* the researcher might conduct a series of regression analyses, or other techniques, to explore which contexts are associated with the various mechanisms—as well as to explore which mechanisms are associated with the various outcomes
* the researcher might examine the association between the mechanisms and outcomes for each context separately
* the researcher might also examine an outcome over time, before and after the intervention, using a technique called interrupted time series analysis (see Nurjono et al., 2018)

The qualitative data, such as answers from interviews, would be subjected to other approaches. To illustrate

* content analysis or thematic analysis could be undertaken to characterize the mechanisms in more detail
* similar techniques could be undertaken to uncover instances in which participants discussed how a specific mechanism culminated in a particular outcome.

Finally, researchers utilize this information to clarify which features of the intervention are effective and in which circumstances. Typically, while interpreting the findings, researchers integrate the insights derived from quantitative data and the insights derived from qualitative data: That is, they triangulate the data during this phase of interpretation (e.g., Underdown, Norwood, & Barlow, 2013).

|  |
| --- |
| **Techniques to improve realistic evaluations** |

When researchers conduct realistic evaluations, they do not rigidly follow a set of procedures. Instead, to apply the principles that underpin realistic evaluations, researchers need to utilize their ingenuity and creativity. The following table offers some guidelines that could help researchers in this endeavour.

|  |  |
| --- | --- |
| Recommendation | Details or examples |
| Often, to measure the context, individuals might derive one categorical variable from a range of specific measures | * For example, to differentiate the contexts, Underdown, Norwood, and Barlow (2013) divided participants into three clusters: low risk, moderate risk, and high risk of psychological problems * This level of risk was derived from six other measures |
| Researchers often generate a table in which each row corresponds to a distinct context. For each row, the researchers indicates the main mechanisms and outcomes that were observed | |  |  |  | | --- | --- | --- | | Context | Mechanisms | Outcomes | | High risk initially; low risk later | M1, M5, M7 | Improved confidence  Improved engagement | | High risk initially; moderate risk later | M1-M4, M9 | Reduced confidence  Reduced engagement | | High risk initially; low risk later | M3, M10 | Improved confidence  Improved engagement | | Moderate risk initially; low risk later | M1, M2, M4-M6 | Improved confidence | | Moderate risk initially and later | M3 | Improved confidence | | Low risk | M7 | Reduced confidence |   NB. M1 to M9 refer to mechanisms that appear in another table |
| Conduct or recommend additional studies | * Sometimes, after interpreting the data, researchers update their context-mechanism-outcome configuration and then test these modified hypotheses again. * Alternatively, researchers might conduct, or recommend that peers conduct, a randomized control trial to assess some of the specific hypotheses more definitively (see Underdown, Norwood, & Barlow, 2013) |
| When conducting realistic evaluations, most researchers adopt a paradigm called critical realism. | * Critical realism accepts the notion that realities are true and objective—and are not merely human constructions—but that researchers cannot depend on scientific reasoning to understand the world * That is, although some realities are objectively true, our knowledge of the world is dependent upon our history, culture and social environment * Researchers can utilize criteria to decide which accounts of the world are more accurate or better |

|  |
| --- |
| **Underlying philosophy** |

Thus far, this document has merely delineated an example of realistic evaluations, designed to demonstrate the aims of this methodology and the main techniques. Yet, to appreciate this methodology, you also need to appreciate the underlying philosophy of realism. The following table compares realism to positivism and constructivism.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Positivism | Constructivism | Realism |
| Ontology: What is the name of our world | The world contains real, tangible objects and patterns that exist regardless of our awareness | Humans construct their interpretations of what the world entails. | Although the world contains real, tangible objects and patterns, we shape this world and the world shapes us reciprocally |
| Epistemology: What is knowledge and how do we acquire knowledge | We strive to uncover truth—that is, insights about the world. To seek this truth, we confine our attention to objects or patterns we can measure objectively | Because every individual constructs their knowledge, we can never ascertain what is real. Arguably, all possible interpretations of reality are equally valid | Although the interpretations of individuals about the world are important, reality does shape which interpretations are reasonable |
| Role of researchers | The researchers measure and observe the world as objectively as possible. So, quantitative research is prioritized | The researchers gather and describe how people interpret the world. So qualitative research is prioritized. | The researchers explain how, why, and when specific interventions or behaviours are effective. So both quantitative and qualitative research is embraced. |
| What is causation | Causation is a sequence of observable events: One event causes a change in some object or system | Humans develop assumptions about the nature of causation—and these assumptions shape our responses to events | Causal forces, called mechanisms, are real but the setting and circumstances affects the operations of these mechanisms |

To help appreciate this table, a few principles should be considered. Specifically

* note that realism is designed to be practical—revolving around when and why various practices are effective—and thus overlaps closely with pragmaticism. However, realism applies more specific assumptions about ontology and epistemology than pragmatism
* realistic evaluations apply the philosophy of realism to evaluations of interventions; realistic reviews applies the philosophy of realism to reviews of the literature or other sources; realistic research entails all research that applies the philosophy of realism
* one common variant, critical realism, also adopt the perspective of less powerful individuals or sectors, exploring issues around justice and power.

|  |
| --- |
| **Alternatives to realistic evaluations** |

To evaluate complex interventions, researchers have adopted and promulgated other approaches as well. For example, the medical research council developed and promoted a framework that evaluates these interventions. This framework comprises a sequence of five distinct phases: theory, modelling, exploratory trial, definitive randomized control trial, and long-term implementation. The following table defines, clarifies, and illustrates these five phases.

|  |  |
| --- | --- |
| Phase | Details |
| **Theory**: Utilize relevant theory to optimize the intervention, to propose hypotheses, to improve the design, and to control relevant variables. | * Extract relevant theories and findings, primarily from past literature, to develop the intervention and to design the study. * Alternatively, you could derive some of these insights, especially about the intervention, from practitioners who already have applied this intervention or similar interventions in the past. |
| **Modelling:** Identify the features of the intervention, as well as the underlying mechanisms, to generate predictions on how these facets relate to each other. | * Researchers can apply a range of designs to clarify the features, or components, of this intervention and articulate the underlying mechanisms * For example, to achieve this goal, they might conduct focus groups, observational studies, preliminary surveys, and computer simulations. |
| **Exploratory trial**: Conduct preliminary studies to assess which features, control groups, outcome measures, and participants you should include in the main study. | * Researchers, for example, might assess the impact of including or excluding various features of this intervention on a range of outcomes. * The researcher might use a variety of methods, such as videos, to gauge which features of the intervention the practitioner applies. In addition, researchers could evaluate various doses or levels of the intervention, diverse control groups, a range of participants, and many outcomes. * For example, the research might assess the feasibility and features of various control groups, such as alternative treatments, placebos, standard care, or other approaches. * These exploratory trials can also clarify the likely effect size—and thus the necessary sample size—the feasibility of outcome measures, and the practices that foster compliance with the intervention. |
| **Definitive randomized control trial.** Compare the optimal intervention to an appropriate alternative, coupled with a defensible protocol, suitable controls, and sufficient statistical power. | * All the features of suitable randomized control trials, such as blind assignment, treatment, and assessment, need to be considered. * For this phase, outcomes should be measured as directly as possible; in contrast, during the exploratory phase, indirect surrogates or predictors of these outcomes might be sufficient. * The intervention should be registered as well. |
| **Long-term surveillance**: Ascertain the enduring effects of this intervention in the natural setting. | * Researchers typically conduct another study to establish whether the intervention is beneficial in a natural setting, outside the controlled environment. * This study also gauges whether these effects are durable and whether adverse events might transpire in the future. * Researchers may also observe the implementation of interventions to identify barriers and other vital insights. |

These five phase, as delineated by the Medical Research Council (2000), do not necessarily contradict the principles of realistic evaluation. Indeed, several authors have argued that researchers may be able to combine realistic evaluation and the five phases (e.g., Craig & Petticrew, 2013; Fletcher, 2016). In general, these authors maintain that researchers should adopt the perspective of realistic evaluations—such as an emphasis on why, when, and how interventions are effective—at each phase of the evaluation.

|  |
| --- |
| **References** |

Byng, R., Norman, I., & Redfern, S. (2005). Using realistic evaluation to evaluate a practice-level intervention to improve primary healthcare for patients with long-term mental illness. Evaluation, 11(1), 69-93.

Cheyne, H., Abhyankar, P., McCourt, C. (2013). Empowering change: realist evaluation

of a Scottish Government programme to support normal birth. Midwifery, 29, 1110-1121.

Clark, A. M., Whelan, H. K., Barbour, R., & MacIntyre, P. D. (2005). A realist study of the mechanisms of cardiac rehabilitation. Journal of Advanced Nursing, 52(4), 362-371.

Craig, P., & Petticrew, M. (2013). Developing and evaluating complex interventions: Reflections on the 2008 guidance. International Journal of Nursing Studies, 50(5), 585–587.

Fletcher, A. (2016). Realist complex intervention science: Applying realist principles across all phases of the Medical Research Council framework for developing and evaluating complex interventions. Evaluation, 22(3), 286-303.

Gill, M., & Turbin, V. (1999). Evaluating “realistic evaluation”: Evidence from a study of CCTV. Crime Prevention Studies, 10, 170–199.

Hewitt, G., Sims, S., &Harris, R. (2012). The realist approach to evaluation research: An introduction. International Journal of Therapy and Rehabilitation, 19(5), 205-259.

Horrocks, I., & Budd, L. (2015). Into the void: A realist evaluation of the eGovernment for You (EGOV4U) project. Evaluation, 21(1), 47-64. doi: 10.1177/1356389014566133

Kirkpatrick, D. L., & Kirkpatrick, J. D. (1994). Evaluating training programs. Berrett-Koehler Publishers

Nurjono, M., Shrestha, P., Lee, A., et al. (2018). Realist evaluation of a complex integrated care programme: Protocol for a mixed methods study. BMJ Open, e017111. doi:10.1136/ bmjopen-2017-017111

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.

Medical Research Council (2000). A framework for development and evaluation of RCTs for complex interventions to improve health.

Pawson, R (2013). The science of evaluation. London: Sage.

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

Porter, S., & O’Halloran, P. (2012). The use and limitation of realistic evaluation as a tool for evidence-based practice: A critical realist perspective. Nursing Inquiry, 19(1), 18-28.

Tolson, D., McIntosh, J., Loftus, L., & Cormie, P. (2007). Developing a managed clinical network in palliative care: A realistic evaluation. International Journal of Nursing Studies, 44(2), 183–95

Underdown, A., Norwood, R., & Barlow, J. (2013). A realist evaluation of the processes and outcomes of infant massage programs. Infant Mental Health Journal, 34(6), 483-495. doi: 10.1002/imhj.21408

Walshe, C., & Luker, K. A. (2010). District nurses’ role in palliative care provision: a realist review. International Journal of Nursing Studies, 47(9), 1167–1183.

Wong, G., Greenhalgh, T., & Westhorp, G., & Pawson, R. (2012). Realist methods in medical education research: What are they and what can they contribute? Medical Education, 46, 89-96.

Wong, E. L. Y., Yeoh, E., Chau, P. Y. K., Yam, C. H. K., Cheung, A. W. L., & Fung, H. (2015). How shall we examine and learn about public-private partnerships (PPPs) in the health sector? Realist evaluation of PPPs in Hong Kong. Social Science & Medicine, 147, 261-269.