

Quality Standards for Realist Evaluation

For evaluators and peer-reviewers

The RAMESES II Project



1. The evaluation purpose

Realist evaluation is a theory-driven approach, rooted in a realist philosophy of science, which emphasises an understanding of causation and how causal mechanisms are shaped and constrained by context. This makes it particularly suitable for evaluations of certain topics and questions – for example, complex interventions and programmes that involve human decisions and actions. A realist evaluation question contains some or all of the elements of ‘What works, how, why, for whom, to what extent and in what circumstances, in what respect and over what duration?’ and applies a realist logic to address the question(s). Above all realist evaluation seeks to answer ‘how’ and ‘why?’ questions. Realist evaluation always seeks to explain. It assumes that programme effectiveness will always be conditional and is oriented to improve understanding of the key contexts and mechanisms contributing to how and why programmes work.

Criterion	Inadequate	Adequate	Good	Excellent
<p>A realist approach is suitable for the purposes of the evaluation.</p> <p>That is, it seeks to improve understanding of the core questions for realist evaluation.</p>	<p>The evaluation does not seek to explain how and why the evaluand* works.</p> <p>or</p> <p>There was no clear statement of the purpose(s) of the evaluation.</p>	<p>The evaluation seeks to explain how and why the evaluand works (or not) and to disaggregate outcomes for different sub-groups and contexts.</p> <p>There is a statement of purpose(s) for the evaluation.</p>	<p>Adequate plus:</p> <p>The evaluation seeks to explain how and why the evaluand works differently in different contexts and for different sub-groups: it seeks to explain how contexts affected mechanisms.</p>	<p>Good plus:</p> <p>Stated purpose clearly explains how the findings are intended to be used. There is a coherent argument as to why a realist approach is appropriate to those purposes.</p>
<p>The evaluation question(s) are framed to be suitable for a realist evaluation.</p>	<p>The evaluation question(s) are not structured to reflect the elements of realist explanation. For example, the question(s):</p> <ul style="list-style-type: none"> • require only description; <i>and/or</i> • require only a numerical aggregation of outcomes; <i>and/or</i> • require only a summary of processes; <i>and/or</i> • rely exclusively on methods that are inadequate to generate realist understanding (e.g. ‘a thematic analysis of ...’) 	<p>The evaluation question(s) include a focus on how and why outcomes were generated in the evaluand, and contained at least some of the additional elements, “for whom, in what contexts, in what respects, to what extent and over what durations”.</p>	<p>Adequate plus:</p> <p>The questions address as many aspects of the realist question as are feasible within the constraints of the evaluation. The rationale for excluding any elements of ‘the realist question’ from the evaluation question(s) is explicit.</p> <p>(For example, the evaluation question may have only sought to explain how and why outcomes occur in certain contexts and not to what extent: the rationale for excluding ‘extent’ is described and reasonable in the circumstances.)</p> <p>The question(s) are sufficiently focused to be managed within the constraints of the evaluation.</p>	<p>Good plus:</p> <p>The evaluation question(s) are clear and as simple as possible. They can be understood by people without specialist methodological or content expertise.</p>

*evaluand: is defined as ‘that which is being evaluated’, for example an intervention, programme, policy, product or initiative, or in some cases, sets of programs, policies or initiatives

2. Understanding and applying a realist principle of generative causation in realist evaluations

Realist evaluations are underpinned by a realist principle of generative causation. That is, underlying causal processes (called 'mechanisms') operate (or not) in certain contexts to generate outcomes. The explanatory framework is Context + Mechanism = Outcome (CMO). Realist evaluations aim to understand how different mechanisms generate different outcomes in different contexts. This intent influences everything from the type of evaluation question(s) to an evaluation's design (e.g. the construction of a realist programme theory, recruitment process and sampling strategy, data collection methods, data analysis, to recommendations).

Criterion	Inadequate	Adequate	Good	Excellent
A realist principle of generative causation is applied.	<p>Significant misunderstandings of realist generative causation are evident. Common examples include:</p> <ul style="list-style-type: none"> • programme/intervention activities or strategies are mislabelled as mechanisms • contexts are assumed to be directly causal, rather than affecting whether and how mechanisms operate • no attempts are made to uncover mechanisms • outcomes are assumed to be caused by the programme/intervention (rather than by underlying mechanisms) • relationship(s) between an outcome, its causal mechanism(s) and context(s) are not explained or configured • if theory is provided, this is not explicitly linked to context-mechanism-outcome configurations 	<p>Some misunderstandings of realist generative causation are evident, but the overall approach is consistent enough that a recognisably realist analysis results from the process.</p>	<p>Assumptions and methods used throughout the evaluation are consistent with a realist generative causation.</p>	<p>Good plus: The evaluation strategy demonstrates exemplary understanding of a principle of realist generative causation, and application of methods consistent with that understanding throughout (for example, in question(s), design and the evaluations outputs). Emerging challenges arising as the evaluation unfolds are dealt with in ways that are consistent with realist generative causation.</p>

3. Constructing and refining a realist programme theory or theories

At an early stage in the evaluation, the main ideas that went into the making of an intervention, programme or policy (the programme theory or theories – which may or may not be realist in nature) are surfaced and made explicit. An initial tentative programme theory (or theories) is constructed which sets out how and why an intervention, programme or policy is thought to ‘work’ to generate the outcome(s) of interest. Where possible, this initial tentative theory (or theories) will be progressively refined over the course of the evaluation. Over the course of the evaluation, if needed, programme theory (or theories) are ‘re-cast’ in realist terms (describing the contexts in which, populations for which, and main mechanisms by which, particular outcomes are, or are expected to be, achieved.) Ideally, the programme theory is articulated in realist terms prior to data collection, in order to guide the selection of data sources about context, mechanism and outcome. However, in some cases, this will not be possible and the product of the evaluation will be an initial realist programme theory.

Criterion	Inadequate	Adequate	Good	Excellent
An initial tentative programme theory (or theories) is identified and developed. Programme theory is ‘re-cast’ and refined as realist programme theory.	<p>Programme theory (or theories) are:</p> <ul style="list-style-type: none"> • not developed; <i>or</i> • not articulated; <i>or</i> • described but not used in the evaluation; <i>or</i> • offered but not ‘re-cast’ and refined as realist programme theory at any stage of the evaluation. In other words, the programme theory is not expressed in terms of the causal relationship between contexts, mechanisms and outcomes. 	<p>Initial tentative programme theory (or theories) are identified and (as far as possible) described in realist terms (that is, in terms of the causal relationship between contexts, mechanisms and outcomes). These are refined as the evaluation progresses.</p> <p>Appropriate data are used to ‘test’ (confirm, refute or refine) selected aspects of programme theory (or theories).</p> <p>Aspects of theory to be ‘tested’ (or not) are:</p> <ul style="list-style-type: none"> • specified and justified in the evaluation design. • appropriate to the purpose of the evaluation. <p>The refined theory (or theories) are consistent with the evidence provided.</p> <p>Basic implications of the final programme theory (or theories) for practice in contexts examined in the evaluation are described.</p>	<p>Adequate plus: Programme theory (or theories) are initially described in realist terms and used to inform all aspects of the evaluation (e.g. focusing an evaluation, identifying questions, determining what types of data are needed from whom and where).</p> <p>A range of appropriate types of data are used to test selected aspects of the theory, including triangulating evidence.</p> <p>Implications of the final programme theory (or theories) for practice in a range of contexts are described. A clear rationale is provided for the contexts in which the findings are applicable or not applicable.</p> <p>Where relevant, the programme theory or theories take into account the physical/ material (e.g. environmental) and social aspects of systems necessary to answer evaluation questions and purposes.</p>	<p>Good plus: The relationships between the programme theory (or theories) and relevant substantive theory (or theories) are articulated.</p> <p>A wide range of primary and secondary data is used to consolidate programme theory.</p> <p>Refinements to substantive theory are described where appropriate.</p> <p>The final realist programme theory (or theories) comprises one or more context-mechanism-outcome configurations, describing how and why different mechanisms are triggered (or not) in different contexts to generate different outcomes.</p> <p>Implications of the final programme theory for a diverse range of contexts are comprehensively described. Relevant contexts which are not included in the evaluation were expressly addressed.</p>

4. Evaluation design

Descriptions and justifications of what is planned in the evaluation design, in what order, and why should be clearly articulated. Realist evaluations are ideally adaptive – that is, the evaluation question(s), scope and/or design may be adapted over the course of the evaluation to ‘test’ (confirm, refute or refine) aspects of the programme theory as it evolves. If changes are made to the evaluation design, these should be clearly described and justified. At the start of an evaluation, where possible, any changes that might be needed should be anticipated and contingencies planned.

Criterion	Inadequate	Adequate	Good	Excellent
The evaluation design is described and justified.	<p>The evaluation design is not clearly described or is not coherent.</p> <p>There is a lack of clarity as to what was planned in the evaluation design, in what order and why.</p> <p>The evaluation design does not clearly relate to or test the program theory. For example, data collection methods used were unlikely to collect the relevant data needed to ‘test’ aspects of programme theory (see the ‘Data collection methods’ section below for more details).</p> <p>Planned analyses are inconsistent with the assumptions underpinning realist evaluation (see the ‘Data analysis’ section below for more details).</p>	<p>What was planned in the evaluation design, in what order and why is described and justified in detail.</p> <p>The evaluation design is informed by initial programme theory or theories, and ‘tests’ important or priority aspects of these.</p> <p><i>or</i></p> <p>The evaluation is appropriately designed to develop realist programme theory.</p> <p>The design is coherent, with a logical flow from purpose through focus, questions, data collection and analysis methods.</p>	<p>Adequate plus: The design ‘tests’ multiple aspects of programme theory.</p> <p>The design enables alternative explanations to be investigated.</p>	<p>Good plus: The design is efficient, adding value by (for example) maximising use of existing data or increasing portability of findings.</p> <p>The design enables consideration of the extent to which the intervention contributes to overall outcomes, and/or identification of other aspects of the context (e.g. other policies or programmes) which are likely to contribute to outcomes.</p>
Ethical clearance is obtained if required.	<p>No consideration is given to whether the evaluation required ethical approval.</p> <p>Ethical approval should have been sought but is not (or was sought and declined).</p>	<p>Protocols for ethics approval are considered and approval sought if required.</p> <p>Where ethical approval is sought, actions throughout the evaluation are consistent with the requirements of the ethics clearance obtained.</p>	<p>Proposals for ethical approval clearly distinguish the implications of the evaluation for different groups and different contexts.</p> <p>The proposal for ethical approval identifies the strategies for iteration in the design and steps to manage ethics in relation to such iteration.</p>	<p>Specific implications of realist methodology are explained in the proposal for ethical approval (e.g. the need to link data across context, mechanism and outcome; the role of the evaluator(s) in relation to other stakeholders and the programme) and specific strategies to address those implications are included.</p>

5. Data collection methods

In a realist evaluation, a broad range of data increases the robustness of the theory 'testing' process and a range of methods used to collect them. Data will be required for all of context, mechanism and outcome and to inform the relationships between them. Data collection methods should be adequate to capture not only intended but also (as far as possible) unintended outcomes (both positive and negative), and the context-mechanism interactions that generated them. Realist evaluation is usually multi-method (i.e. uses more than one method to gather data). Where possible, data about outcomes should be triangulated (at least using different sources, if not different types, of information).

Criterion	Inadequate	Adequate	Good	Excellent
Data collection methods are suitable for capturing the data needed in a realist evaluation.	<p>Within the realist evaluation project:</p> <ul style="list-style-type: none"> • data collection methods are unclear <p><i>and/or</i></p> <ul style="list-style-type: none"> • data collection methods are not theory driven (i.e. informed by the need to find data to confirm, refute or refine the programme theory) <p><i>and/or</i></p> <ul style="list-style-type: none"> • methods used are unlikely to capture necessary data (i.e. all of context, mechanism and outcome and the relationships between them). 	<p>Methods for collecting and documenting data are driven by the programme theory (or theories) and:</p> <ul style="list-style-type: none"> • capture the necessary data, including sampling necessary to test the programme theory; <p><i>and</i></p> <ul style="list-style-type: none"> • capture intended and unintended outcomes. <p><i>and</i></p> <ul style="list-style-type: none"> • address the evaluation questions. <p>The rationale for the methods used is explained.</p>	<p>Adequate plus:</p> <p>Data collection methods are explicitly consistent with realist methodology (e.g. realist interviewing).</p> <p>Quality control processes are adopted to ensure that data collection methods are applied rigorously and consistently.</p> <p>Allowance is made to collect additional data for further refinement of programme theory (or theories) and/or CMO configurations as the evaluation unfolds.</p> <p>Data management processes (e.g. data bases, use of participant identifiers) are constructed to enable intended analyses (e.g. sub-group analyses, tracking participants over time).</p>	<p>New data collection methods, tools and processes are adapted and/or developed where required and are consistent with realist principles.</p> <p>The specific techniques used or adaptations made to instruments or sampling processes are justified.</p>

6. Sample recruitment strategy

In a realist evaluation data are required for all of C, M and O. One key source is respondents or key informants. Data are used to develop and refine theory about how, for whom, and in what circumstances programmes generate their outcomes. This implies that any processes used to invite or recruit individuals need to identify an adequate sample of individuals who are able to provide information about contexts, mechanisms, outcomes and/or programme theory.

Criterion	Inadequate	Adequate	Good	Excellent
The respondents or key informants recruited are able to provide sufficient data needed for a realist evaluation.	<p>Recruitment is not designed to find respondents who could provide information about contexts, mechanisms and/or outcomes (e.g. recruitment was ad hoc and/or not informed by the programme theory (or theories).</p> <p>Random samples are used to generalise to whole populations (as distinct from sampling within theory-specified sub-groups).</p> <p>Convenience samples are used to 'test' (as distinct from develop) programme theories.</p>	<p>Recruitment is:</p> <ul style="list-style-type: none"> designed to find an appropriate sample of respondents who can provide information about contexts, mechanisms and/or outcomes and the programme theory. purposive, with samples selected to test specific aspects of programme theory. 	<p>Adequate plus:</p> <p>Where needed, further recruitment is undertaken to collect the data needed for refinement of programme theory and/or CMO configurations.</p>	<p>Sampling follows a rigorous and sequenced process of theory testing.</p> <p>A sufficiently large and diverse sample of relevant respondents is recruited to provide evidence across contexts and sub-groups.</p> <p>When needed, respondents are re-interviewed as new evidence emerges, to explore context and mechanism extensively.</p> <p>Where applicable, sampling involves sensitive strategies to successfully recruit respondents from disenfranchised communities or other 'hard to reach' groups.</p>

7. Data analysis

Data analysis in realist evaluation is not a specific method but a way of interrogating programme theory (or theories) with data and a way of using theory to understand patterns in data. In other words, data analysis is a way of teasing out what works, for whom, in what contexts, in what respects, over what duration and so on.

In a realist evaluation, where possible, the analysis process should occur iteratively. The overall approach to data analysis is retroductive* (i.e. it moves between inductive and deductive processes, includes and tests researcher ‘hunches’ and aims to provide the best possible explanation of acknowledged-to-be-incomplete data). The processes used to analyse the data and integrate them into one or more realist programme theories should be consistent with a central principle of realism - namely generative causation. How these data are then used to further develop, confirm, refute or refine one or more programme theories should be clearly described and justified.

Criterion	Inadequate	Adequate	Good	Excellent
The overall approach to analysis is retroductive*.	The approach to analysis is not retroductive. <i>or:</i> The overall approach to analysis is not clear.	The approach to analysis moves between theory and data, data and theory appropriate to the stage of theory development.	Adequate plus: Theory is developed and refined through the use of retroductive reasoning. Evaluators’ ‘hunches’ are clearly described. Theories that remain untested are specified and described.	Good plus: Analysis clearly links data, programme theory and formal theory.
Data analyses processes applied to gathered data are consistent with a realist principle of generative causation.	Analytic processes are not described. Analysis is not disaggregated by sub-groups (i.e. ‘for whom’) and/or contexts. Subgroup analyses are undertaken without reference to programme theory (for example, disaggregating by gender, age or other demographic sub-groups without specifying how they are relevant to theory, rather than on theory-relevant groupings).	Qualitative analysis moves beyond thematic categorisation to identify and explain the relationships between contexts, mechanisms and outcomes. Quantitative analysis is hypothesis driven to ‘test’ differences between sub-groups or contexts, in relation to programme theory. Findings from analysis are organised to demonstrate relationships between context, mechanism and outcome (i.e. evidence is aligned against programme theory).	Adequate plus: Specific analyses are conducted to ‘test’ the relationships within and between CMOCs (e.g. correlations analysis for quantitative data; analysis of narrative, argument or speech/text to identify causal relationships in qualitative data). That is, evidence is not just aligned against programme theory: the linkages within the programme theory are ‘tested’. Weaknesses in analytic methods for realist purposes were acknowledged and choices justified.	Good plus: When iterations in evaluation design and/or programme theory required additional analytic methods, these methods are consistent with realist principles.

7. Data analysis continued

Criterion	Inadequate	Adequate	Good	Excellent
A realist logic of analysis is applied to develop and refine theory.	<p>The analyses does not:</p> <ul style="list-style-type: none"> • identify contexts, mechanisms or outcomes; • identify the relationships between contexts, mechanisms and outcomes <p><i>and/or</i></p> <ul style="list-style-type: none"> • describe how the programme theory (or theories) was further developed, confirmed, refuted and refined 	<p>Data are analysed to develop and refine initial programme theory (or theories) into realist programme theory (or theories).</p> <p>The realist analysis:</p> <ul style="list-style-type: none"> • assigns conceptual labels of C, M or O to each data element or finding within a Context-Mechanism-Outcome configuration (CMOC) – (e.g. ‘in this aspect of the analysis, this item of data is functioning as context’). • identifies the relationship between contexts, mechanisms and outcomes within particular CMOCs. • identifies relationships across CMOCs – i.e. the location and interactions between CMOCs within a programme theory (or theories) 	<p>Adequate plus:</p> <p>Analysis:</p> <ul style="list-style-type: none"> • integrates a range of data sources (e.g. qualitative and quantitative, primary and secondary data) <i>and</i> • describes how multiple data types were integrated to support inferences. 	<p>Data analysis is iterative over the course of the evaluation, with earlier stages of analysis being used to refine programme theory and/or refine evaluation design for subsequent stages.</p>

* For more details on retrodution see: ‘Retrodution in realist evaluation’ which may be found in the Standards and Training materials section of The RAMESES Projects website (www.ramesesproject.org).

8. Reporting

Realist evaluations may be reported in multiple formats – detailed reports, summary reports, articles, websites and so on. Reports should be consistent with the RAMESES II reporting standards for realist evaluations (see <https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-016-0643-1>).

Criterion	Inadequate	Adequate	Good	Excellent
The evaluation is reported using the items listed in the RAMESES II reporting standard for realist evaluations.	<p>Key items are missing. For example:</p> <ul style="list-style-type: none"> • No defined evaluation question(s) • Limited or no reporting of the evaluation's methods • Limited or no explanations and justifications provided for any adaptations made to the realist evaluation approach • Insufficient detail to enable readers to judge the trustworthiness and plausibility of findings. 	<p>Most items in the RAMESES II reporting standards for realist evaluations are reported. In particular:</p> <ul style="list-style-type: none"> • Item 3: Rationale for evaluation • Item 4: Programme theory • Item 5: Evaluation questions(s), objectives and focus • Item 6: Ethics • Method section items 8 (Environment for the evaluation), 9 (Description of the evaluand), 11 (Data collection methods), 12 (Recruitment and sampling), 13 (Data analysis) and 15 (Main findings). 	<p>All items are clearly reported and in sufficient detail for an external reader to understand and judge the methods used and the trustworthiness and plausibility of findings.</p> <p>Where an item is not reported, a justification is provided.</p>	<p>Good plus:</p> <p>Additional materials are made available for external readers to investigate aspects of the evaluation in more detail.</p>
Findings and implications are clear and reported in formats that are consistent with realist assumptions.	<p>Findings are unclear or difficult to follow.</p> <p>Findings are not reported in realist format (for example, average results are reported but do not address issues such as 'for whom' or 'in what circumstances').</p> <p>Lists of contexts, mechanisms and outcomes are provided without reporting causal relationships between them.</p> <p>Evidence is not clearly linked to context, mechanism or outcome.</p>	<p>Findings are clearly reported.</p> <p>All conclusions follow logically from the analyses.</p> <p>Findings explain how and why different patterns of outcomes are generated in different contexts or for different groups.</p>	<p>Adequate plus:</p> <p>Implications for policy, programmes and/or practice are clearly explained and follow logically from the analysis.</p> <p>Implications and/or recommendations take into account issues or strategies for different contexts or groups.</p> <p>Summaries of findings maintain patterns of outcomes (e.g. findings are not summarised by resorting to average effects).</p>	<p>Good plus:</p> <p>The report is well written, transparent, and easy to understand.</p> <p>Various reporting formats are used to present relevant findings to different audiences.</p>

The RAMESES II Project

Funder: NIHR HS&DR

Project team:

Professor Trish Greenhalgh UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/trish-greenhalgh

Professor Ray Pawson UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/pawson

Dr Geoff Wong UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/geoffrey-wong

Dr Gill Westhorp CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/greenhalgh

Dr Ana Manzano UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/manzano

Dr Justin Jagosh UNIVERSITY OF LIVERPOOL
www.liverpool.ac.uk/psychology-health-and-society/staff/justin-jagosh

This project was funded by the National Institute of Health Research Health Services and Delivery Research Programme (project number 14/19/19).

Professor Trish Greenhalgh's salary is part-funded by the Oxford Biomedical Research Centre, NIHR grant number BRC-1215-20008.

The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the HS&DR programme, NIHR, NHS or the Department of Health.

For information on how these quality standards were developed and how to use them please see the full project report which may be found on The RAMESES Project website (www.ramesesproject.org).



NUFFIELD DEPARTMENT OF
PRIMARY CARE
HEALTH SCIENCES
Medical Sciences Division



NHS
*National Institute for
Health Research*