

Observing Leadership: Using Non-participant Observation to Support Programme Theory Development in Realist Evaluation

International Journal of Qualitative Methods

Volume 25: 1–8

© The Author(s) 2026

DOI: 10.1177/16094069261480503

journals.sagepub.com/home/ijqLisa Knight¹ 

Abstract

Non-participant observation is a recognised method in qualitative research but remains underutilised in realist evaluation, with most published examples focused on frontline service delivery settings. This paper offers a methodological reflection on the specific contribution of non-participant observation to programme theory development within a realist evaluation of leadership and governance in an Integrated Care System (ICS) in England. The article argues that governance settings, where leadership is enacted through discourse, negotiation, and formal decision-making, present distinctive methodological challenges that make observation particularly valuable.

Drawing on eleven observations of governance meetings, conducted alongside interviews and documentary analysis, this paper demonstrates how non-participant observation contributed to the iterative development and refinement of context-mechanism-outcome (CMO) configurations. Observation provided access to interactional and relational processes that participant accounts did not fully capture, including non-verbal cues, patterns of participation, and divergences between espoused and enacted practice. Where interview accounts described collaborative decision-making, observations of the same forums sometimes revealed more hierarchical dynamics, with these divergences becoming productive starting points for retroductive reasoning. Fieldnotes also informed the design of subsequent interviews, enabling specific observed incidents to be used as prompts to explore participant reasoning, consistent with realist interviewing principles.

Limitations of observation in governance settings, including restricted access to informal interactions, interpretive uncertainty, and the constraints imposed by the formality of meeting structures, are presented. The paper concludes that non-participant observation can function as an active component of early theory development in realist designs involving complex organisational settings, and calls for greater methodological transparency in how observation is used and reported within realist evaluation.

Keywords

realist evaluation, non-participant observation, programme theory, context-mechanism-outcome, leadership, integrated care, qualitative methods

Received: 28 March 2026; Revised: 3 August 2026; Accepted: 7 August 2026

¹Liverpool John Moores University, Liverpool, Merseyside

Corresponding author:

Lisa Knight, Liverpool John Moores University, Liverpool L2 2QP, Merseyside.

Email: l.knight@ljmu.ac.uk



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Introduction

Realist evaluation offers a theory-driven approach to understanding what works, for whom, in what circumstances, and why (Pawson & Tilley, 1997). Grounded in a critical realist ontology, it seeks causal explanation: rather than asking only whether an intervention works, it asks how and why interventions generate outcomes through the activation of mechanisms within particular contexts. Central to this approach is the identification and refinement of programme theories, often expressed as context–mechanism–outcome configurations (CMOCs). These configurations are developed through iterative engagement with evidence, using data collection methods that can illuminate the interaction between contexts, mechanisms, and outcomes. Within these configurations, mechanisms are commonly understood as the interaction between the resources or opportunities an intervention introduces and the reasoning and responses of those who encounter them (Dalkin et al., 2015). Interviews, focus groups, and documentary analysis are frequently reported in realist evaluation as the primary means of generating insights into reasoning and response (Manzano, 2016; Renmans & Castellano Pleguezuelo, 2022), with comparatively few studies incorporating observation or other methods capable of capturing behaviour. Participant accounts, however carefully elicited, are retrospective and bounded by what actors are willing and able to articulate.

This paper offers a methodological reflection on the use of non-participant observation to support programme theory development and refinement within a realist evaluation of leadership and governance in healthcare. The evaluation explored how leadership practices contributed to system-level objectives and outcomes, within an Integrated Care System (ICS), where leadership operates across organisational boundaries through multi-actor governance structures. ICSs are statutory multi-organisational systems, established in England in 2022, that bring together NHS provider trusts, primary care, local authorities, and voluntary and community organisations to commission and deliver integrated health and social care across a defined geography. The evaluation focused on a single ICS and examined how leadership worked across partner organisations operating within shared governance structures; specifically, how collaborative decision-making, shared vision, and inter-organisational relationships were enacted in practice. Governance meetings were the primary site through which these leadership activities were observable, making them the natural focus for non-participant observation within the study design.

The study utilised observations, interviews, and documentary analysis to inform the development and refinement of programme theories. The focus in this paper is not the findings of the evaluation itself, but the role that observation played in supporting theory development and refinement within this methodological approach. In reflecting on this experience, the paper seeks to contribute to ongoing methodological discussions about data collection in realist evaluation and to support evaluators considering the use of observation in studies. Specifically, the paper argues that in governance settings observation is most valuable not as a supplementary or triangulating method but as a methodologically distinct source of evidence: its divergences from participant accounts, particularly between espoused and enacted practice, provide productive starting points for retroductive reasoning and the refinement of programme theory. Key outcomes of interest in the evaluation included the quality and consistency of inter-organisational decision-making, the degree of collaborative engagement across partner organisations, and the alignment, or misalignment, between strategic intent and operational practice within the ICS governance structure.

Observation in a Realist Evaluation

Realist evaluation is underpinned by the principle that social programmes work (or not) through the activation of mechanisms within specific contexts to generate outcomes (Pawson & Tilley, 1997). Programme theories, often articulated as context–mechanism–outcome configurations (CMOCs), are developed, tested, and refined through iterative engagement with multiple forms of evidence (Emmel et al., 2018). Throughout, programme theory refers to the broad explanatory account of how an intervention is expected to work; a CMOC is its formalised expression as a context–mechanism–outcome configuration; the mechanism is the generative component within that configuration; and theory refinement denotes the iterative revision of these accounts as evidence accumulates. Data collection strategies are typically selected to support this theory-driven process, to elicit participant reasoning and contextual insights (Manzano, 2016). Although interviews remain central to realist approaches, constituting the dominant method across published realist evaluations (Renmans & Castellano Pleguezuelo, 2022), observation methods have been identified as valuable for examining practices as they occur, generating empirical evidence relevant to the theorisation of mechanisms or contextual conditions that are difficult to access through self-report (Handley et al., 2020; Marchal et al., 2010).

However, the limited published examples offered tend to focus on the use of observation in frontline service delivery settings, where mechanisms are often linked to interpersonal behaviours, task performance, or environmental cues (Handley et al., 2020). There remains comparatively limited reflection on the use of observation in governance and organisational

settings, where mechanisms may operate through more abstract processes such as alignment, influence, or negotiation, and where action is often mediated through language, agenda-setting, and formal decision-making procedures (Nicolini, 2012).

Realist theorising must attend simultaneously to individual reasoning, relational dynamics, and system-wide patterns (Shearn et al., 2017). Whereas in frontline service-delivery settings mechanisms are frequently linked to relatively proximal and observable phenomena, such as interpersonal behaviours, task performance, or responses to environmental cues (Handley et al., 2020), in governance settings mechanisms operate at a greater remove from individual action such as alignment, framing, deference, and agenda-setting, rather than discrete tasks (Jarzabkowski et al., 2007; Nicolini, 2012). Against this backdrop, observation provides access to the enacted conditions through which mechanisms may plausibly operate. This includes how power, deference, and influence are performed through discourse, silence, non-verbal cues, and agenda management in ways that participants may not consciously articulate or may retrospectively obscure. Observation can also document observable outcomes, such as whether decisions are reached collaboratively or dominated by particular actors, alongside the contextual conditions shaping them, reinforcing its value as a multi-functional tool in realist designs (Renmans & Castellano Pleguezuelo, 2022).

The realist evaluation that informs this paper focused on understanding leadership, with particular attention to identifying the mechanisms through which leadership activities influence outcomes. A multi-method, theory-driven design was employed, combining documentary analysis, interviews, and non-participant observation. Initial programme theories were developed through a literature review and exploratory data collection, comprising interviews and observations, and were then refined iteratively across the study phases, consistent with realist practice (Emmel et al., 2018; Knight et al., 2024; Rycroft-Malone et al., 2012).

Observation is recognised as a way of capturing naturally occurring practices and interactions, offering a means of generating empirical material from which to theorise about mechanisms as they operate in real time (Handley et al., 2020). Utilising Gold's (1958) classic continuum of observer roles as a conceptual grounding, from complete participant to complete observer, the described study was situated at the non-participant end of that continuum, with the researcher refraining from active involvement while remaining attentive to naturally occurring behaviour. In realist evaluation, mechanisms are understood as latent, generative forces rather than directly observable phenomena (Mukumbang, 2023; Pawson & Tilley, 1997). Observations do not therefore surface mechanisms directly but provide the empirical conditions under which mechanism theorisation can take place.

In studies where interventions are implemented through direct service delivery, observation has been used to explore behaviours, responses, and environmental influences that may not be fully articulated in interviews or documents (Handley et al., 2020; Marchal et al., 2010). In such settings, the relationship between observable action and underlying mechanism is often foregrounded, with observation contributing to the process of testing or refining programme theories. An advantage of observation over interviews alone is its capacity to cross-check subjective actor accounts against enacted behaviour: actors may not always be sufficiently reflexive to articulate the mechanisms shaping their own practice, or may describe idealised versions of practice that differ from what actually occurs (Handley et al., 2020). Less attention has been given in the realist evaluation literature to the use of observation in leadership and governance practices. In such settings, mechanisms operate through discourse, negotiation, and decision-making processes rather than observable task performance alone, and formal processes such as meetings, agenda-setting, and strategy development both enact and constrain leadership in ways that are difficult to access through self-report (Jarzabkowski et al., 2007; Nicolini, 2012). Moreover, innovative approaches to data collection in realist evaluation, including observation, remain underexplored across the field more broadly, with qualitative interviews continuing to dominate as the primary, and often sole, data collection tool (Renmans & Castellano Pleguezuelo, 2022).

Study Design and Observational Approach

Within this design, observation was used to support programme theory development and refinement, rather than solely for testing pre-specified theories. The approach drew on Manzano's (2016) notion of theory gleaning, recognising the potential for observation to contribute to the early identification of candidate mechanisms and contextual conditions. A total of eleven observations were conducted across two phases of data collection over a six-month period. Permission to attend meetings was negotiated in advance with ICS governance leads, although meetings were held in public. In the first phase, a largely unstructured approach was adopted, with the researcher attending meetings and documenting whatever was salient to the emerging programme theories without a predetermined focus. In the second phase, a semi-structured topic guide, informed by, but not restricted to, the initial programme theories developed in phase one, was used to direct attention towards specific contextual factors, interactional patterns, and potential mechanisms. A semi-structured approach to observation was adopted consistent with recommendations from the broader qualitative methods literature, whilst balancing attention to emerging theoretical concerns against the openness that unstructured observation affords (Kawulich, 2005; Mulhall, 2003). Beginning

without a predetermined focus was important given the provisional nature of the emerging initial programme theories. As these developed, more targeted attention became possible, with observation directed towards contextual conditions and interactional patterns that had emerged as significant. This progression from open to focused observation supported a deepening of analytic sensitivity as the researcher developed familiarity with the setting (Mulhall, 2003).

Observational data were recorded by hand during meetings, to avoid the disruption associated with laptop use in small and formal settings. Notes taken during sessions were necessarily condensed, capturing verbatim exchanges where possible and interactional summaries where real-time transcription was not feasible. These were expanded into full fieldnotes immediately after each session, while memory of the interaction remained fresh. Fieldnotes recorded not only the content of discussion but also the structure of participation: who spoke, in what sequence, who was deferred to, for example. Attention to these features was deliberate, given the argument that in governance settings, the organisation of discourse is itself a form of action (Nicolini, 2012). Decisions about what to record were inevitably selective, and this selectivity was documented through reflexive journaling, enabling the researcher to return to questions of interpretive emphasis during analysis.

Fieldnotes were coded in ATLAS.ti using an initial framework derived from the candidate programme theories, supplemented by open codes for interactional features such as turn-taking, deference, silence, and agenda control. Observational and interview data were analysed together rather than as separate strands, so that observed sequences and participant reasoning could be read against one another. Candidate mechanisms were identified where recurrent interactional patterns suggested a plausible generative explanation, and configurations were recorded in a working context–mechanism–outcome table that was revised as evidence accumulated, drawing on Dalkin et al.'s (2015) separation of mechanism resources and reasoning. Where observations and interviews diverged, for instance where power dynamics observed in meetings contradicted interview descriptions of collaborative decision-making, the divergence was retained as an analytic prompt rather than resolved in favour of either source. It became a starting point for retroductive reasoning, working backwards from the observed pattern to theorise which underlying mechanisms, operating under which contextual conditions, could plausibly have generated the outcome (Mukumbang, 2023; Pawson & Tilley, 1997).

In this way, observational data supported or challenged interview-derived mechanisms and actively generated new explanatory hypotheses that were tested through subsequent interviews. This iterative relationship between observation and interviews enabled the evaluation to examine how observed behaviours and interactions related to the reasoning processes described by participants, supporting the ongoing refinement of CMOCs.

To illustrate how observation contributed to theory refinement, a worked example is offered: an initial programme theory suggested that, in the context of bringing together health and social care partners with different histories and governance arrangements, the development of strong relationships through shared goals and frequent interaction (mechanism resource) would nurture trust and a sense of shared ownership (mechanism reasoning), supporting collaborative, jointly-owned decision-making (outcome). Interview accounts of governance forums frequently described decision-making in these collaborative terms.

Observation of the same forums, however, sometimes recorded more hierarchical patterns, particularly under conditions of financial pressure. Discussion often narrowed around the priorities of the largest or statutorily accountable organisations, and whilst contributions from smaller partners were acknowledged they were seldom incorporated, and partners were observed to orient towards organisational rather than system interests when resources were contested. This divergence between the interviewee accounts of collaboration and the enacted interaction became a starting point for retroductive reasoning focused around the question: what contextual conditions would have to be present for a shared vision to give way, in practice, to more siloed and centralised decision-making?

Working backwards from the observed pattern, the candidate mechanism was reconfigured: rather than partner engagement straightforwardly producing collaboration, financial pressure and system fragmentation (context) appeared to activate organisational self-protection and a reassertion of hierarchy (mechanism), producing decisions that were collaborative in form but centralised in substance (outcome). The observed sequences were then used as prompts in subsequent interviews, inviting participants to reason about why retrenchment occurred under financial strain. In this way an observed contradiction, rather than confirming the initial theory, helped to reshape it.

Reflections on the Use of Observation for Programme Theory Development

One of the key benefits of observation was its potential to provide insight into discursive, relational, and interactional processes through which leadership activity was performed. Observation enabled attention to be given to how alignment, negotiation, and the establishment of shared purpose occurred through discursive interaction (Golsorkhi et al., 2015), supporting the identification of relational processes as candidate mechanisms, and providing a complementary perspective to the reasoning accounts elicited through interviews. A particular advantage of observation in this governance setting was its capacity to surface aspects of leadership that actors themselves may not fully articulate, for the reasons discussed above,

providing an independent empirical basis for identifying gaps between espoused and enacted practice. Fieldnotes from observations informed the design of interviews, enabling specific incidents or interactions to be used as prompts to explore participants' reasoning about leadership actions and decision-making processes. This practice aligns with [Manzano's \(2016\)](#) guidance on realist interviewing, where the integration of prior knowledge and observation can enhance the quality of data collected by grounding questions in concrete examples. This approach also reflects wider debates in realist evaluation regarding the importance of integrating diverse data sources to support the identification of CMOs ([Wong et al., 2013](#)).

On several occasions, interview accounts described governance processes in terms of collaborative decision-making and shared ownership, while observations of the same forums revealed more hierarchical patterns. As the worked example above illustrates, such divergences were not merely noted but became an analytical pivot for retroductive reasoning with the observed interaction, read against the espoused account, generating the question from which a refined configuration was developed. Therefore, the ability of observation to capture non-verbal cues, interactional dynamics, and environmental features further contributed to the development of programme theories. Observation can offer access to data that remain implicit in verbal accounts, including body language, silences, spatial arrangements, and contextual conditions ([Weston et al., 2021](#)). Recent methodological guidance emphasises the importance of observation for generating insight into the complexities of healthcare settings, particularly where understanding processes and interactions in context is critical ([Weston et al., 2021](#)). In this evaluation, such features were particularly relevant in formal leadership meetings, where expressions of agreement, resistance, or deference were sometimes conveyed through these non-verbal means. This dimension of observation enriched the data available for theory refinement, supporting a more nuanced understanding of how mechanisms operated within these settings.

However, the use of observation also raised questions regarding interpretation, positionality, and researcher reflexivity. Observing formal governance spaces involves navigating issues of access, visibility, and power, which shape what can be observed and how observational data are understood ([Yip, 2024](#)). Reflexivity was embedded throughout the study as an active analytical practice, consistent with calls to move beyond performative positionality statements towards deeper critical engagement with how researcher subjectivity shapes data collection and interpretation ([Downey et al., 2024](#)).

In the observed settings, the position of the researcher as an external observer within hierarchical forums required ongoing consideration of the ways in which presence may have influenced meeting dynamics, and of how interpretations of interaction were shaped by the researcher's background knowledge and experience. The researcher approached these settings from a background of leadership in healthcare settings, bringing prior familiarity with how boards and partnership forums are expected to operate. This familiarity drew attention to features such as chairing behaviour, patterns of deference, and the management of dissent, but also carried the risk of reading meetings through an insider frame, for instance interpreting silence as deference when it might equally signal disengagement, tactical withholding, or settled prior agreement. These assumptions were treated as data to be interrogated rather than as neutral starting points and were tested through discussion of alternative readings within the research team. Accordingly, reflexive journaling was maintained throughout the study to document interpretive shifts and analytical decisions, providing an audit trail of how observational data were integrated into theory development.

Challenges and Limitations of Observation in Leadership and Governance Practices

While non-participant observation provided valuable support for programme theory development in this realist evaluation, several challenges and limitations were encountered. A primary limitation concerned the 'observability' of mechanisms. Observational methods rely on visible actions, discourse, and interactional patterns, which may not fully capture the cognitive or emotional dimensions of mechanism activation ([Emmel et al., 2018](#)). Although observation supported the identification of interactional and relational mechanisms, the underlying reasoning of participants often required elaboration through interviews.

A more specific interpretive risk involves the move from observed interaction to theorised mechanism with an observer attuned to one reading of an interaction potentially over-inferring from it. This risk can be minimised by holding inferences as provisional, testing them against participant reasoning in subsequent interviews. Related to this was the issue of interpretive uncertainty. Observation in formal meetings and governance settings involves interpreting what is seen and heard, including the significance of silences, non-verbal cues, and interactional dynamics. These interpretations are unavoidably shaped by the researcher's background knowledge, assumptions, and positionality. Systematic patterns in what the observer notices, and what passes unnoticed, can shape programme theory development in ways that are difficult to identify after the fact. Reflexive journaling and discussion of interpretations within the research team offers a mitigation to this, but the risk cannot be wholly eliminated.

Access represented a further constraint; observation was limited to meetings and forums where permission was granted, which meant that certain informal or closed discussions were not observed. This reflects a broader issue in organisational

research, where gatekeeping and selective access shape what can be studied and how comprehensive observational data can be (Delamont, 2004). Unpacking gatekeeping in institutional settings involves navigating a dynamic social web of actors and relationships rather than a single point of entry or refusal (Spencer et al., 2023). In realist terms, this may limit the ability to observe the full range of contexts or mechanisms relevant to programme theory development. The observations conducted provided insight into the operation of mechanisms within formal spaces, but mechanisms activated in informal, private, or interpersonal interactions remained less accessible.

Ethical dimensions of observation in governance settings also require attention, though they are infrequently discussed in methodological accounts of realist evaluation. Formal governance meetings of public sector bodies, as described in this reflection, are typically open by statute or convention, and observation of such forums often does not require individual consent in the same way that observation of private or clinical interactions would. Nevertheless, the presence of a researcher whose purpose is to generate data about the behaviour of identifiable actors raises ethical considerations. In this study, ethical approval was obtained through institutional processes, and the study was noted with the ICS as a research activity. The tension between naturalistic observation and transparency was managed by making the purpose of the researcher's presence explicit to participants at the outset of each meeting, rather than relying on the public character of the forum as a substitute for informed engagement.

The structure and formality of governance meetings also shaped what could be observed. Meetings followed established protocols, including agendas, minute-taking, and formal turn-taking, which structured participation and potentially constrained the expression of disagreement or alternative views. While these features of context were analytically significant, they also limited the scope for spontaneous interaction or informal reasoning to be observed. This aligns with reflections that formal processes may mask underlying tensions or negotiations that occur outside official forums (Roberts et al., 2005).

Finally, there were challenges associated with temporal limitations and continuity of observation. Non-participant observation provides snapshots of practice rather than continuous engagement, and decisions about which meetings to observe and when to attend influenced what was captured. This raises questions about the representativeness of observational data and the potential for key mechanisms to be missed or misunderstood if they occurred outside observed sessions. While the iterative integration of observation with interviews and document analysis mitigated some of these risks, the episodic nature of observation remained a limitation.

These challenges highlight the importance of approaching observation in realist evaluation with methodological caution and reflexivity. Observation can make a valuable contribution to programme theory development, particularly in contexts where mechanisms are enacted through discourse and relational practices, but its interpretive limitations, ethical considerations, and contextual constraints require careful attention.

Conclusion and Implications for Realist Evaluation Practice

The reflections presented highlight three key methodological features of observation within realist evaluation. First, observation can support the early stages of theory development by surfacing candidate mechanisms and contextual conditions that may not be fully accessible through participant accounts alone. This extends previous methodological discussions that have tended to position observation as a tool for theory testing or triangulation. Empirical evidence from qualitative studies confirms that prolonged observational engagement produces contradictions with interview data, revealing divergences between participants' stated practices and their enacted behaviour (Dado et al., 2023). This makes observation particularly valuable not as a supplementary method but as a methodologically distinct source of evidence that can actively challenge and refine programme theories. Second, observation contributes to a richer understanding of how mechanisms are enacted in real-time interactions, including non-verbal cues, patterns of participation, and environmental conditions that shape behaviour and reasoning. Third, observation can enhance the integration of data sources by informing interview design and supporting iterative engagement with evidence across methods, consistent with realist principles of theory refinement (Manzano, 2016; Rycroft-Malone et al., 2012).

However, the methodological challenges encountered in this evaluation also emphasise the importance of approaching observation with reflexivity and caution. Mechanisms may remain partially or wholly inaccessible to observation alone, requiring integration with interviews and other methods to develop robust programme theories. The interpretation of observed events is shaped by the positionality of the observer, and the structured nature of some settings may constrain both what can be seen and what can be said. These limitations reinforce the need for critical engagement with how observational data are generated, interpreted, and integrated within realist designs. As Downey et al. (2024) argue, reflexivity remains underutilised in realist evaluation more broadly; the use of observations creates additional occasions for reflexive practice and should be accompanied by explicit documentation of how the researcher's positionality shaped what was visible, what was interpreted, and what was prioritised in analytical decisions.

The reflections offered here point to an underutilised contribution that non-participant observation can make within realist evaluation, particularly in settings where mechanisms are enacted through discourse and relational practice rather than observable task performance. Used alongside interviews and documentary analysis, observation provides access to the enacted dimensions of practice that participant accounts cannot fully capture, and the divergences it surfaces between espoused and enacted behaviour generate productive starting points for retroductive reasoning and programme theory refinement that other methods in the design may not reach. An open-to-focused progression, in which early unstructured observation serves theory-gleaning purposes before giving way to more directed attention as initial programme theories develop, fits the iterative logic of realist design and works most effectively when treated as a planned methodological commitment from the outset. There is, however, a case for greater consistency in how observation is reported across realist evaluations: how access and sampling were negotiated, how observational and interview data were analytically integrated, and how the researcher's positionality shaped what was visible and how it was interpreted are dimensions that remain underspecified in much of the published literature. Greater transparency on these points would both strengthen the credibility of observation as a component of realist designs and support evaluators in making well-founded decisions about its use in complex organisational contexts.

ORCID iD

Lisa Knight  <https://orcid.org/0000-0002-5095-8896>

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

- Dado, M., Spence, J. R., & Elliot, J. (2023). The case of contradictions: How prolonged engagement, reflexive journaling, and observations can contradict qualitative methods. *International Journal of Qualitative Methods*, 22, 16094069231189372. <https://doi.org/10.1177/16094069231189372>
- 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(1), 49. <https://doi.org/10.1186/s13012-015-0237-x>
- Delamont, S. (2004). Ethnography and participant observation. In C. Seale, G. Gobo, J. F. Gubrium, & D. Silverman (Eds.), *Qualitative Research Practice* (pp. 217–229): Sage.
- Downey, J., Harris, N., Rybczynska-Bunt, S., Golder, E., Shearn, K., & Bradley, N. (2024). Reflecting on reflexivity in realist evaluation: a call to action. *International Journal of Qualitative Methods*, 23, 16094069241284206. <https://doi.org/10.1177/16094069241284206>
- Emmel, N., Monaghan, M., Manzano, A., & Greenhalgh, J. (2018). *Doing realist research*: Sage.
- Gold, R. L. (1958). Roles in sociological field observations. *Social Forces*, 36(3), 217–223. <https://doi.org/10.2307/2573808>
- Golsorkhi, D., Rouleau, L., Seidl, D., & Vaara, E. (2015). Introduction: What is strategy as practice?. In *Cambridge Handbook of Strategy as Practice* (2nd ed., pp. 1–30). Cambridge University Press.
- Handley, M., Bunn, F., Lynch, J., & Goodman, C. (2020). Using non-participant observation to uncover mechanisms: Insights from a realist evaluation. *Evaluation*, 26(3), 380–393. <https://doi.org/10.1177/1356389019869036>
- Jarzabkowski, P., Balogun, J., & Seidl, D. (2007). Strategizing: The challenges of a practice perspective. *Human Relations*, 60(1), 5–27. <https://doi.org/10.1177/0018726707075703>
- Kawulich, B. B. (2005). Participant observation as a data collection method. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(2), 43. <https://www.qualitative-research.net/index.php/fqs/article/view/466>
- Knight, L., Neiva Ganga, R., Tucker, M. P., Shore, A. P., & Nolan, S. (2024). Contexts and complexities: A realist evaluation of integrated care system leadership. *Leadership in Health Services*, 38(1), 153–169. <https://doi.org/10.1108/LHS-06-2024-0051>
- Manzano, A. (2016). The craft of interviewing in realist evaluation. *Evaluation*, 22(3), 342–360. <https://doi.org/10.1177/1356389016638615>
- Marchal, B., Dedzo, M., & Kegels, G. (2010). A realist evaluation of the management of a well-performing regional hospital in Ghana. *BMC health services research*, 10(1), 24. <https://doi.org/10.1186/1472-6963-10-24>

- Mukumbang, F. C. (2023). Retroductive theorizing: a contribution of critical realism to mixed methods research. *Journal of Mixed Methods Research*, 17(1), 93–114. <https://doi.org/10.1177/15586898211049847>
- Mulhall, A. (2003). In the field: Notes on observation in qualitative research. *Journal of Advanced Nursing*, 41(3), 306–313. <https://doi.org/10.1046/j.1365-2648.2003.02514.x>
- Nicolini, D. (2012). *Practice theory, work, and organization: An introduction*. Oxford University Press.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage.
- Renmans, D., & Castellano Pleguezuelo, V. (2022). Methods in realist evaluation: A mapping review. *Evaluation and Program Planning*, 97, 102209. <https://doi.org/10.1016/j.evalprogplan.2022.102209>
- Roberts, J., McNulty, T., & Stiles, P. (2005). Beyond agency conceptions of the work of the non-executive director: Creating accountability in the boardroom. *British Journal of Management*, 16(s1), S5–S26. <https://doi.org/10.1111/j.1467-8551.2005.00444.x>
- Rycroft-Malone, J., McCormack, B., Hutchinson, A. M., DeCorby, K., Bucknall, T. K., Kent, B., Schultz, A., Snelgrove-Clarke, E., Stetler, C. B., Titler, M., Wallin, L., & Wilson, V. (2012). Realist synthesis: Illustrating the method for implementation research. *Implementation Science*, 7, 33. <https://doi.org/10.1186/1748-5908-7-33>
- Shearn, K., Allmark, P., Piercy, H., & Hirst, J. (2017). Building realist program theory for large complex and messy interventions. *International Journal of Qualitative Methods*, 16, 1–11. <https://doi.org/10.1177/1609406917741796>
- Spencer, K. L., Mrig, E. H., & Bouchard, E. G. (2023). Unpacking gatekeeping in medical institutions: A case study of access to end-of-life patients. *Qualitative Research*, 23(2), 486–500. <https://doi.org/10.1177/14687941211034975>
- Weston, L. E., Krein, S. L., & Harrod, M. (2021). Using observation to better understand the healthcare context. *Qualitative research in medicine & healthcare*, 5(3), 9821. <https://doi.org/10.4081/qrmh.2021.9821>
- Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J., & Pawson, R. (2013). RAMESES publication standards: Realist syntheses. *BMC Medicine*, 11, 21. <https://doi.org/10.1186/1741-7015-11-21>
- Yip, S. Y. (2024). Positionality and reflexivity: negotiating insider-outsider positions within and across cultures. *International journal of research & method in education*, 47(3), 222–232. <https://doi.org/10.1080/1743727x.2023.2266375>