Understanding the process of a research utilisation intervention in policy agencies: What, how and why?

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Abby Haynes

Sydney School of Public Health
Faculty of Medicine
University of Sydney

2017
Statement of authenticity and authorship attribution

To the best of my knowledge, the work presented in this thesis is original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for another degree at this or any other institution.

I certify that the intellectual content of this thesis is the product of my own work. I led the conceptual and practical work including: designing and leading the research, collecting the data, conducting the data analyses, and drafting the included publications. I further declare that all assistance received in preparing this thesis and sources has been acknowledged. Contributions from others are as follows:

The process evaluation contributors—Sue Brennan, Gisselle Gallego, Stacy Carter and Denise O’Connor—provided conceptual and operational advice during the development and conduct of the process evaluation via regular meetings and periodic workshops. They contributed to data analysis and writing up findings.

Chapters Four, Six, Seven, Eight and Nine contain manuscripts that have been published, or accepted for publication, in peer-reviewed journals. The process evaluation contributors and some members of the investigator team commented on draft manuscripts. My supervisors—Phyllis Butow, Sally Redman and Anna Williamson—provided guidance throughout. Specific input to publications was as follows:

**Chapter Four** - *What can we learn from interventions that aim to increase policymakers’ capacity to use research? A realist scoping review.*

To strengthen the rigour of this review, Samantha Rowbotham independently read the included studies, developed hypotheses and worked with me to iteratively critique and refine the findings, and develop ways of expressing them. Prior to this, Sue Brennan read the synthesis of studies and critiqued my initial context+mechanism=outcome configurations.

**Chapter Seven** - *Figuring out fidelity: a worked example of the methods used to identify, critique and revise the essential elements of a contextualised intervention in health policy agencies.*

Emma Darsana helped test the fidelity constructs by attending several intervention workshops and independently completing fidelity coding sheets for comparison with my coding.

**Chapter Nine** - *Policymakers’ experience of a capacity building intervention designed to increase their use of research: a realist process evaluation.*

Sue Brennan independently analysed a proportion of the data and worked with me to reconcile and articulate our findings. Steve Makkar provided frequency calculations (Table 1 and Figure 1 in Appendix 14) and Sue Brennan and Joanne McKenzie commented on the statistics. Emma Darsana helped with confirmatory data collection and coding.

All other co-authors of publications included in this thesis—Tari Turner, Andrew Milat, Gabriel Moore and Sian Rudge—were involved in critically revising the manuscript they are listed on.

I should also clarify where my contribution ends. I was not an investigator on the grant that funded SPIRIT. I was involved in early work on sampling and in developing the constructs that informed the study, but I did not participate in the final design or implementation of the intervention or the measures.

Megan Howe proofread this thesis, providing feedback on typos, formatting and unclear language.
Abby Haynes

Signature: 

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Supervisor name: Phyllis Butow

Signature: 

# Table of contents

Part 1. Context and background ........................................................................... 6

Abstract .................................................................................................................. 7

Chapter One: Introduction and overview ................................................................. 8
   Intervening to increase the use of research ............................................................ 9
   Research rationale, aims and objectives ............................................................... 9
   Outline of chapters ............................................................................................ 11

Chapter Two: Theory and methodology ................................................................. 13
   Epistemology and ontology ............................................................................... 13
   Theoretical perspective ..................................................................................... 15
   Methodological approach ............................................................................... 16
   Research quality and credibility ....................................................................... 18

Chapter Three: The challenge of increasing research use in policymaking ............... 22
   Why intervene in policy agencies’ use of research? ............................................. 22
   Challenges to improving the use of research in policymaking .............................. 23
   The need for research-informed policy ............................................................... 31
   Getting research into policy ............................................................................ 31

Chapter Four: A review of research utilisation interventions in policy agencies .......... 33
   Overview .......................................................................................................... 33
   Manuscript ....................................................................................................... 34

Part 2. Framing the intervention and process evaluation ......................................... 65

Chapter Five: The SPIRIT study .......................................................................... 66
   Supporting Policy In health with Research: an Intervention Trial ......................... 66
   The SPIRIT process evaluation ....................................................................... 70

Chapter Six: Defining SPIRIT’s core concepts ...................................................... 79
   Overview .......................................................................................................... 79
   Manuscript ....................................................................................................... 80


Chapter Seven: Flexibility and fidelity ................................................................. 92
   Overview .......................................................................................................... 92
   Manuscript ....................................................................................................... 95
Chapter Eight: The role of internal facilitators ................................................................. 117
  Overview.......................................................................................................................... 117
  Manuscript...................................................................................................................... 118

Chapter Nine: Participants’ experience of SPIRIT .......................................................... 138
  Overview.......................................................................................................................... 138
  Manuscript...................................................................................................................... 139

Chapter Ten: Discussion ................................................................................................. 166
  Understanding what happened in SPIRIT, and why ....................................................... 166
  Contribution to methods in this area ............................................................................. 170
  Contribution to the development of theory in this area ............................................... 171
  Future research directions ............................................................................................ 174

Thesis references ............................................................................................................. 177

Appendices ....................................................................................................................... 199
  Appendix 1. Glossary and acronyms ........................................................................... 200
  Appendix 2. Models of research utilisation ................................................................. 204
  Appendix 3. Review characteristics and search strategies ......................................... 210
  Appendix 4. Overview of included studies .................................................................. 213
  Appendix 5. Context and process .............................................................................. 231
  Appendix 6. Early interview questions (general participants) ...................................... 233
  Appendix 7. Post-intervention interview questions (general participants) ............... 235
  Appendix 8. Heads used in the Framework Analysis matrices .................................... 237
  Appendix 9. Chapter Six manuscript .......................................................................... 238
  Appendix 10. Chapter Seven Manuscript ................................................................... 254
  Appendix 11. An example of how selected essential elements changed during SPIRIT .................................................................................................................. 273
  Appendix 12. Chapter Eight Manuscript ................................................................... 275
  Appendix 13. Chapter Nine Manuscript ..................................................................... 303
  Appendix 14. Descriptive overview of results ............................................................. 324
  Appendix 15. Supporting theory for the realist process evaluation analysis .................. 332
  Appendix 16. Publications arising from this thesis ...................................................... 338
  Appendix 17. Acknowledgements .............................................................................. 340
Part 1.
Context and background
Abstract

Background
There are increasing calls to strengthen the use of research in health policymaking, but little is known about how intervention strategies may be received by policymakers, or how different contexts are likely to affect engagement and uptake. This research seeks to understand, and provide transferable information about, how research utilisation interventions function in different policy settings, and how they can be made more fit for purpose.

Approach
This thesis focuses on Supporting Policy In health with Research: an Intervention Trial (SPIRIT). SPIRIT is a multi-component research utilisation intervention that was implemented in six health policy agencies in Sydney. Taking a process-orientated perspective, my mixed methods research examines facets of the intervention design and implementation, and attempts to both describe and explain how a range of participants perceived and interacted with SPIRIT, and with what effects.

Work produced
Six studies are included. Five published papers: (i) a realist scoping review of interventions aimed at building policymakers’ capacity to use research in their work; (ii) a participant observation study of concept development in our team; (iii) an account of theory-focused fidelity assessment; (iv) an analysis of the views, behaviours and impacts of the intervention’s internal facilitators; and (v) a realist evaluation of how participants experienced SPIRIT and the causal pathways through which intervention strategies appear to generate process effects. A further chapter explores the wider context and considerable challenges of increasing research use in policymaking.

Key contributions
This thesis delivers new conceptual and methodological approaches for understanding how and why complex interventions function as they do. First, the findings describe how SPIRIT was implemented and perceived, and offer provisional explanations for the marked variation in engagement between the six intervention sites. Using an in-depth realist process evaluation approach, I was able to identify and test possible causal mechanisms, and to make empirically grounded recommendations for program improvement (e.g. improved strategies for identifying and supporting internal facilitators).

Second, the thesis makes methodological contributions. It advances the use of realist process evaluation, which is still rarely used, unpacks the role of process effects in a realist scoping review of research utilisation interventions, and presents a novel approach for managing fidelity assessment within flexible interventions. The inclusion of pragmatic tools and worked examples makes this work concrete.

Lastly, the research contributes to theory in the field. It furthers our understanding of the dynamic and highly situated connections between policymakers’ diverse information needs and practices, and different kinds of research utilisation intervention and implementation strategies. It brings this learning together in provisional transferable propositions and provides access to the substantial empirical and conceptual work that sits behind them. Together, these contributions offer guidance for the design, implementation and evaluation of future intervention studies in this field and beyond.

The thesis concludes by highlighting areas where further understanding is required if the ambitions of research-informed policymaking are to be advanced.
Chapter One: Introduction and overview

“Too often we ignore the how of change”
Zimmerman et al. [1:10]

Interventions are hard to understand. What goes on during implementation? How do activities designed to bring about change actually function? And why do participants so often experience the ‘same’ intervention in vastly different ways? The answers to such questions are often cloaked in a metaphorical black box: we know what goes in and what comes out, but relatively little about what happens in between. The interactions and mechanisms that generate change are a mystery. [2, 3] Without sound information about what happens in this black box we cannot make informed judgements about program improvement or know which parts of the intervention should be delivered in other settings, including how to upscale and tailor it for local needs. [4] This information is also required for interpreting the study outcomes, determining the extent to which the findings are transferable, and for differentiating between implementation failure and inadequate program theory. [5-8] In short, understanding how and why interventions work—not simply whether they work—is critical. [2]

Interventions that seek to change professional practices within organisations are legendarily challenging. [9-12] Experts in this field point to the many studies that fail to have an impact and argue that, all too often, the design, implementation and evaluation of interventions does not take adequate account of fundamental considerations about process and context. [13, 14] This means we not only fail to understand our own intervention studies, we also fail to provide useful empirical information that could guide others. [15] There are two central and overlapping areas of inquiry that demand attention:

First, what are the differences between how an intervention is envisaged and what actually happens in each setting? It is seldom possible (or, many argue, desirable) to implement the same suite of activities in precisely the same manner across diverse settings, circumstances and participants [2, 15]. Rather, the form and function of an intervention is actively shaped by the people who participate in it and the circumstances in which it is delivered. [5, 11, 16] These interactions are especially unpredictable when interventions are new (in their first iteration) [2], comprised of multiple components, and designed to allow local tailoring. [17-19].

Second, how do people targeted by the intervention perceive and interact with its activities and resources, and how is this shaped by context? Context here means “the set of circumstances or unique factors that surround a particular implementation effort”. [13] This includes participants’ perceptions of the intervention and of their work practices; intra-organisational culture and relationships; meso and macro systems and structures; shifting trajectories of change; and the attitudes and behaviours of those involved in delivering or otherwise mediating the intervention in each site. [2, 5, 13, 20, 21] A critical aspect of this inquiry is to develop and/or build on theories so that findings relate not only to the circumstances from which they derived, but can be transferred meaningfully to other studies and settings. [15, 22, 23]

These questions are of interest in studies of standardised interventions where the intention is to maximise adherence, but equally to those trialling flexible interventions where the aim is to tailor locally appropriate strategies. In both cases there will be variation within and between sites. This is a feature of implementation and of inevitable contextual heterogeneity. A growing body of evaluators argue that a locus of scientific advancement depends on investigating this variation to build theory that moves us from a-contextual notions of ‘what works?’ and instead develops more nuanced and actionable findings about ‘what works for whom under what circumstances, and why?’. [15, 23]
Increasing recognition of the value in obtaining this deeper, more contextually-grounded and theoretically sound understanding of how interventions function has led to a shift in process evaluation design. [24] The traditional focus on numerical data—how much was delivered to how many people over what periods of time—is expanding to incorporate methods that can provide a rich account of what the intervention comprises when delivered; how its activities are linked to outcomes; and how program, participants and context interact. [5, 25, 26] As we shall see, the use of qualitative methods, and of theory-driven approaches, are pivotal to this endeavour. [17, 27]

Intervening to increase the use of research

Close attention to process and context is particularly pertinent for interventions that aim to increase the use of research in professional practice. Such interventions are likely to be regarded quite differently depending on participants’ roles, skills, beliefs and values, and the organisational culture in which they work. [28] Not least because the use of research in each setting will vary as its ideas and implications for action intersect with tacit experiential knowledge, institutional norms, diverse forms of competing information, stakeholder interests, and cross-disciplinary decision-making processes. [11, 12, 29-31]

Research utilisation interventions in health policy agencies are no exception. Policy agencies are viewed as complex systems in which “myriad elements of context—including different professional, organisational and sectoral cultures and the role of power and politics—are critical considerations.” [32] In these settings, research and other forms of information have to ‘perform’ strategically [33, 34] within policymaking processes that are described as “a contested arena of negotiation… messy, complex, and serendipitous”. [35] Macro-level political and institutional factors influence how policymakers and policy organisations engage with and make use of research [33], and therefore mediate their relationships with research utilisation interventions. Given that the use of research is cultural and rhetorical as well as technical [36], where an intervention promotes greater use of research, or claims to be evidence based, participants may actively critique that premise. [37, 38] Thus, it is vital to determine how such an intervention interacts with participants’ beliefs, practice norms and organisational culture.

Research rationale, aims and objectives

This thesis seeks to improve understanding of how and why interventions work by investigating the key differences between how an intervention is envisaged and what actually happens in each setting, how people targeted by an intervention perceive and interact with its activities and resources, and how these are shaped by context.

My overarching aim was to develop new conceptual and methodological approaches that could guide the design, implementation and evaluation of future intervention studies. I hoped to achieve this through detailed analysis of the functioning of a specific complex research utilisation intervention trial known as Supporting Policy In health with Research: an Intervention Trial (SPIRIT). SPIRIT was a multicomponent cluster randomised trial in six health policy agencies based in Sydney, Australia (described in Chapter Five). Using multiple case study methodology and realist approaches, I sought to produce transferable knowledge that could inform other interventions in this field and beyond.

I designed and led the in-depth process evaluation which enabled me to use a range of research strategies to look at what the intervention comprised and how it was implemented, how the trial was perceived and experienced by different groups of participants in each of the participating agencies, and to develop provisional explanations about these interactions, including why there was such pronounced variation between the intervention sites.
To explore how and why SPIRIT functioned as it did, I examined four different but critical aspects of the trial:
a. The key concepts that underpinned SPIRIT, b. The tensions between intervention flexibility and fidelity, c. The activities and impact of internal facilitators, and d. The perceptions and experiences of participants.

There are three objectives:
1. To understand what happened in SPIRIT, and why
2. To contribute to methods in this area
3. To contribute to the development of theory in this area.
Outline of chapters

<table>
<thead>
<tr>
<th>Part 1. Context and background</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chapter One</strong></td>
</tr>
<tr>
<td>Introduction and overview</td>
</tr>
<tr>
<td><strong>Chapter Two</strong></td>
</tr>
<tr>
<td>Theory and methodology</td>
</tr>
<tr>
<td><strong>Chapter Three</strong></td>
</tr>
<tr>
<td>The challenge of increasing research use in policymaking</td>
</tr>
<tr>
<td><strong>Chapter Four</strong></td>
</tr>
<tr>
<td>A review of research utilisation interventions in policy agencies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part 2. Framing the intervention and process evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chapter Five</strong></td>
</tr>
<tr>
<td>Research context: The SPIRIT study</td>
</tr>
<tr>
<td><strong>Chapter Six</strong></td>
</tr>
<tr>
<td>Defining SPIRIT’s core concepts</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part 3. What happened, and why</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chapter Seven</strong></td>
</tr>
<tr>
<td>Flexibility and fidelity</td>
</tr>
<tr>
<td><strong>Chapter Eight</strong></td>
</tr>
<tr>
<td>The role of internal facilitators</td>
</tr>
<tr>
<td><strong>Chapter Nine</strong></td>
</tr>
<tr>
<td>Participants’ experience of SPIRIT</td>
</tr>
<tr>
<td><strong>Chapter Ten</strong></td>
</tr>
<tr>
<td>Discussion and conclusion</td>
</tr>
</tbody>
</table>
A note about terminology

This thesis draws on concepts, and thus terminology, from several different fields. It also uses some terms quite precisely to reflect specific interpretations, especially where they are contested or poorly defined in the broader literature (e.g. words like research, evidence and context). Although terms are usually outlined when they are first mentioned, the glossary and list of acronyms in Appendix 1 may be helpful.
Chapter Two: Theory and methodology

“... it is obviously unfashionable to suggest that the aim of research methods is to provide some sort of approximation to what is ‘really’ going on. Yet this is, I think, what drives and should drive most social scientists, just as most of us live our everyday lives as though reality exists and can be known about”

Oakley (1999) [39:252]

Crotty [40] identifies four elements that comprise the “scaffold” of assumptions and decision-making that underpins all research. These are: epistemology and ontology, theoretical perspective, methodology and methods. I use this scaffold to put my thesis in its philosophical and theoretical context.

Epistemology and ontology

Epistemology—the nature of knowledge and belief—is the overarching paradigm that directs and substantiates research’s knowledge claims. It provides “a philosophical grounding for deciding what kinds of knowledge are possible and how we can ensure they are both adequate and legitimate”. [40:8] Thus epistemology can be thought of as a justification of knowledge. [41] This includes how research validity is determined. [42] Epistemology is closely linked to ontology—the nature of being and reality—because what we can know is boundaried by our understanding of what is. However, as I will argue below, ontology does not determine epistemology.

Social constructionism

Using Crotty’s definitions, this thesis has a broadly social constructionist¹ approach in assuming that knowledge “is contingent upon human practices, being constructed in and out of interaction between human being and their world, and developed and transmitted within an essentially social context” [40:42]

According to Burr [43], key tenets of social constructionism include:

- Questioning taken-for-granted knowledge. A stance that “invites us to be critical of the idea that our observations of the world unproblematically yield its nature to us, [and] to challenge the view that conventional knowledge is based on objective, unbiased observation”. [43:2]
- Knowledge as a system of shared beliefs created and sustained through social practices. We cannot directly perceive reality so ‘truth’ is only ever our current accepted version of how we understand phenomena.

Constructionism embraces a degree of relativism in that it recognises that “What is said to be ‘the way things are’ is really just ‘the sense we make of them’”. [40] Crucially, however, while meaning does not exist independently outside of our experience, most of the phenomena about which we construct meaning most certainly do.

¹ The terms social constructionism and social constructivism are often used with little distinction, but constructivism is more often associated with a radical individualistic epistemology rather than the stance I take here which focuses on intersubjectivity and the collective development of knowledge (see Crotty, 1998). Consequently, I use social constructionism in this thesis. Where constructivism appears in a quote (e.g. Maxwell on page 6), the author has not made this distinction and is describing what I am calling constructionism.
Realism

This thesis is also strongly informed by a realist paradigm. Philosophically, realism can be viewed as a ‘middle way’ in that it straddles the positivist/constructionist dichotomy by asserting that there is a factual reality beyond our perceptions, but that our knowledge of this reality is partial and interpretive. [44, 45] Thus social phenomena², such as those targeted in process evaluation, are real [46] but our understanding of them is constrained by the limits of socially constructed knowledge and will always be a work-in-progress. [45]

This distinction between ontology and epistemology is critical for understanding realism and how it fits (or can be fitted) with a constructionist approach. Because of our tendency to conflate ontology and epistemology, the ‘real’ in realism is often assumed to include epistemology and thus relegates realism to the objectivist camp where universal truths can be discovered. Meanwhile, because constructionism emphasises human-generated meaning, it is often misunderstood as having a radical subjectivist ontology (i.e. reality exists only within our minds), which would preclude any transferability of constructionist research findings and, arguably, render all research meaningless. In fact, these are both misrepresentations. [40] Using examples as diverse as quarks, madness and Captain Cook, Hacking points out that when we talk about something being socially constructed it is not the thing itself, but our ideas and beliefs about it that are constructions. The object is real, but the subject—i.e. the meaning attributed to the object—is a construct. [47] Crotty argues that this means, “constructionism in epistemology is perfectly compatible with a realism in ontology”. [40:63] Maxwell, a realist, agrees,

“... realists thus retain an ontological realism (there is a real world that exists independently of our perceptions, theories, and constructions) while accepting a form of epistemological constructivism and relativism (our understanding of this world is inevitably a construction from our own perspectives and standpoint).” [45:5]

In both constructionist and realist research, phenomena do not have intrinsic meaning, and findings are not discovered but are theorised by the researcher using an interpretive process informed by the best information available. So realist and constructionists both attend to data with “exactness”. [40:48] The result is not objective truth or subjective belief, but more or less useful interpretations that get closer to reality as researchers triangulate high quality empirical data, and theorise with imagination and insight. [23, 45, 46, 48]

Where realists diverge is their emphasis on testing hypotheses about how interventions are supposed to work. These hypotheses are identified via program theory and through data analysis, and they take the form of propositions that describe how intervention activities bring about effects by triggering responses among certain groups of people in particular circumstances. Propositions, therefore, are causal explanations that contain information about pertinent contextual features, underlying mechanisms and intervention effects, otherwise known as Context+Mechanism=Outcome (CMO) configurations. [15, 23] Unlike positivist hypothesis-testing which seeks universal truths and tends to assume that the conduct of program evaluation should be standardised within and across studies, realists contend that “nothing works unconditionally in all circumstances” [49:126], thus purposive sampling and data collection is used to follow rich veins of information pragmatically as theories take shape and are refined. [48] This means that the status of the data differs somewhat in emphasis. For instance, realists agree with constructionists that data obtained from interviews, observations and document analysis is a construction, but argue that it is also evidence of real phenomena that can be tested, and about which transferable inferences can be drawn. [45]

---

² Entities can be real in different ways. For example, they may be material things, ideas or concepts, artefacts or social phenomena. See Fleetwood, S., Ontology in organization and management studies: a critical realist perspective. *Organization*, 2005. 12(2):197-222.
Theoretical perspective

The theoretical stance or worldview that underpins research provides a rationale for its design and for judging its quality. Making this explicit goes some way towards revealing the “complexus of assumptions” that inevitably inform all inquiry. [40:66]

This thesis is informed by a theoretical perspective that is interpretivist in a loosely Weberian mode; by which I mean that it subscribes to “a science which attempts the interpretive understanding of social action in order to arrive at a causal explanation of its courses and effects”. [50] Arguably, this is a sociological approach in that it contends that social phenomena can only be understood within their social and historical setting, and that this ‘big picture’ encompasses interconnected levels of structure and activity. [51] Further, it attempts to explain as well as describe social phenomena, so grounds theorising in empirical verification. [40] This has much in common with the mode of scientific realism espoused by Pawson [15], Pawson and Tilley [23], and others in the RAMESES II project [52] where the aim of evaluation is to identify “what works, for whom, in what circumstances, and how”. [23, 53]

Some of the overlapping bodies of work that guided my thinking about structure, activity and causality are now outlined. Evidence of their influence as sensitising concepts [54] will be seen repeatedly throughout this thesis.

Sensitising concepts

At the macro level—which is concerned with institutional systems such as government, media and the economy—ideas from systems thinking helped explain how the nested relationship between organisations and wider institutional systems contributes to the mercurial nature of organisational change [32, 55, 56]; and the complexity of policymaking, including the notion of wicked problems [57-59] (which realists reframe in more helpful terms [15]).

In considering the goals of SPIRIT, the debate around evidence-based policy (EBP) was central. EBP is hugely contested: for some a clarion cry, for others a naïve myth. [60] Critics refute the implication that policy can or should be founded on research findings [61, 62]; however, in rejecting the purist conceptualisation of EBP it is vital that we do not throw out the baby with the bathwater. Reliable, fit for purpose research can usefully inform policymaking and efforts to support and enhance this project are worthy. This debate is explored in Chapter Three.

At the meso level—which is concerned with group relationships and culture within disciplines, organisations and communities—the large and diverse body of work on organisational learning and change offered important insights into how and why workplace interventions function as they do. Key influencing ideas included: organisations that engage in reflexive practice are more adept and open to change; organisational structure and systems shape the way people interact with information; leaders model the organisation’s learning climate; ‘resistance’ to change usually has a local rationale that must be addressed; and change often occurs at multiple levels and along different dimensions. [31, 63-66] Systems thinking operated at this meso level too, helping me to contextualise and redefine complex interventions, providing an ecological lens on implementation and invaluable ideas about how fidelity assessment can cope with complexity. [1, 18, 67-71]

Implementation science and the literature that explores the how interventions are perceived helped me recognise that implementation is a contingent process, that interventions are interpreted and enacted differently in different settings, and that context and interpersonal interactions are fundamental evaluation
targets. [10-13, 37, 72, 73] This shaped decisions about what forms of data would best inform my exploration of the relationship between process and outcomes.

Research on collective workplace practices was also valuable; for example, showing how communities of practice shape professional identity and the kinds of knowledge that are valued. Together with the sensemaking literature (below), this helped me reconceptualise knowledge transfer as a social and interpretive process of knowledge transformation in which the development of professional practices relies on surfacing and sharing the tacit knowledge of trusted peers. [31, 74-77] When considering how policymakers engage in these forms of interpersonal knowledge legitimisation, Caplan’s [78] work offered insights. His contention that researchers and policymakers inhabit different worlds with conflicting values, languages, timelines and rewards is overgeneralised [79], but it provides a useful touchstone for articulating some of the core obstacles for research-informed policy. [80] Caplan’s calls for greater collaboration, for researchers to contribute to policy more effectively, and for syntheses of research with policy expertise, remain pertinent.

At the micro level—which is concerned with individual roles, experiences, identities and practices—social psychological learning theories offered insights into how people interact with information. For example, March and Olsen’s models of learning in ambiguous situations [81], and Weick’s Sensemaking theory [82], posit that individual behaviour within organisations is interpretive and retrospective, driven by situations and beliefs as well as attributes and roles, and that people’s interaction with information involves trial and error, chance and superstitious learning.

Bringing many of the above concepts together, the literature exploring how policymakers use research [e.g. 28, 83-87] sensitised me to the gulf between EBP ideals and real world messiness. Portraits of policymaking as “discourse” [88], “rhetorical action” [34, 89], “juggling” [90], “muddling through” [91], an “interplay of ideas” [92] and “primeval soup” [93] contrast sharply with depictions of rational decision-making within a defined policy cycle. This literature problematised normative assertions about what ‘should’ or could happen in terms of practice change, and raised questions about how information and ideas actually get into, and move around within, policy agencies. [85, 94, 95] It also made me appreciate the value of a multidisciplinary perspective [96] that includes political science. [33, 90]

Much of this literature is informed by the concept of indeterminacy: social phenomena tend to be stochastic—creating their own conditions unpredictably—and are often resistant to rational management. Social research must take account of this tendency. [97]

Methodological approach

In this thesis I seek to understand the processes through which an intervention interacts with people and context, and to identify underlying causal patterns in variation between the different settings. My task, therefore was first to explore the breadth of intervention-related perceptions of participants against the backdrop of their different organisational cultures and circumstances, and to develop early hypotheses about how and why these interactions came. Second, I needed to gather data that could help to develop, nuance or refute my evolving hypotheses. Consequently, I used a multiple case study design in which the unit of analysis was the interaction of the intervention with people and features in the organisational setting, i.e. the process of SPIRIT in each site.

According to Yin [98], case studies are most suitable when researching contemporary, complex, social phenomena that are not clearly theorised, and where context is an integral aspect of the phenomena. Knowledge utilisation within an organisation is ripe for case study research given that it has undefined
boundaries, outcomes cannot be reduced to single factor theories, interaction and interdependencies are key, and situations are shaped by aspects such as agency culture and the historical moment in time. [99]

The purpose of a case study approach was to provide comparative description of the interaction between intervention implementation and context in each agency, but also to contribute to theory building about possible causal mechanisms. Flyvbjerg [100] [101] argues that a key advantage of the case study is that it can ‘close in’ on complex situations and test emergent hypotheses directly in relation to phenomena as they unfold in real time. Their focus on process tracing (linking causes and outcomes), exploring possible causal mechanisms, examining the sensitivity of concepts in historical context, and reformulating hypotheses and research questions based on falsification, make case studies extremely valuable for theory building. While not generalisable in the statistical sense, they contribute to the development of theories via “analytic generalisability” where propositions are hypothesised and tested. [98] The rich description of phenomena within case studies increases transferability, i.e. the ability of others to determine how the findings are likely to apply in their contexts. [102]

Six organisations participated in the intervention trial and I was able to work across all of them, enabling me to conduct multiple case studies. The use of multiple cases enables hypotheses derived in one site to be examined in others, expanding opportunities for falsification (e.g. by asking ‘How is the plausible explanation I developed in relation to site A supported, nuanced or disconfirmed by what I have found in sites B and C?). [100] The resulting conclusions have resilience across all the studies cases and are, therefore, more reliable and have greater analytic generalisability. This method is particularly suited to research where ‘How’ and ‘why’ questions are posed about the processes or outcomes of an intervention. [98, 103]

Methods

Given the epistemology, theoretical framework and methodology outlined above, and the purpose of my work (to investigate questions of how and why), my research was mixed-method with an emphasis on qualitative methods. Qualitative research is best positioned to capture “richness, depth, nuance, context, multidimensionality and complexity…. this means that it has an unrivalled capacity to constitute compelling arguments about how things work in particular contexts.” [104:1]

Multiple methods were used in data collection, management and analysis. Data collection included purposively sampled interviews, fieldwork observations, fidelity coding, participant feedback forms, document review, and informal ad hoc conversations with intervention providers and participants. Data analysis included comparative case study analysis, framework analysis, coding for context-mechanism-outcome configurations in the realist mode, and some descriptive statistical analyses. The triangulation of multiple data sources is a cornerstone of case study research and realist evaluation. [41, 48, 52, 98]

Details about data collection and analysis are provided in Chapter Five.
Research quality and credibility

Research should adhere to quality criteria that are appropriate to its theoretical basis, aims and design. [105, 106] Given that this thesis derives from a predominantly qualitative approach, I will focus on issues pertaining to the quality and conduct of qualitative research.

Patton argues, and many others concur, that “The qualitative researcher has an obligation to be methodical in reporting sufficient details of data collection and the processes of analysis to permit others to judge the quality of the resulting product”. [106:1191] But what exactly should be reported? Quality criteria for qualitative research have been the subject of dispute for decades. This dispute is an off-shoot of the “paradigm wars” [39] between scholars with qualitative or quantitative orientations who are, essentially, wrestling with different philosophical worldviews. [107] The paradigm wars themselves are outdated and do not warrant discussion here; suffice to say that the supposed opposition between qualitative and quantitative research is a false dichotomy [e.g. 39, 108-110]. These methods of inquiry are complementary and in many cases work symbiotically to reach understandings that cannot be reached without the other. Further, neither qualitative nor quantitative research are unified fields. [111] The task is to match methods appropriately to the research question rather than assume methods have inherent value, “Just as form follows function in architecture, design follows purpose in research and evaluation”. [108]

However, echoes of the paradigm war and ensuing uncertainty about qualitative research appraisal can be detected in the dispute about which of over 100 sets of qualitative research criteria should be used [105, 111]. Should we apply positivist constructs of reliability, validity and objectivity (referred to by some as a “holy trinity” [105])? Or redefine these constructs to reflect different purposes (a post-positivist approach [102])? Or reformulate our conceptualisation of quality entirely in accordance with the philosophical underpinnings of the research? [102]?

Given that quality appraisal in qualitative research is strongly related to the theoretical perspective in which the study is located [111], and that this thesis includes realist and social constructionist perspectives, it seems appropriate to use quality criteria that straddle post-positivism in general and constructionism in particular. Consequently, I take a ‘middle ground’ and describe my approach to quality using post-positivist criteria that reinterpret positivist constructs, but flesh these out with constructs that were developed to strengthen the credibility of constructionist research. [106, 112]

Morrow [102] syntheses work by Lincoln and Guba [113, 114] and Gasson [115] to identify four post-positivist criteria for assessing qualitative research: credibility, transferability, dependability and confirmability. I use these criteria to explain how I ensured research quality throughout this study (Table 2.1). The strengths and weaknesses of the approaches used in substudies are discussed in each relevant chapter.

---

3 The fourth option of doing away with quality criteria altogether is given little attention.
# Table 2.1. Quality criteria and how I addressed them in my research

<table>
<thead>
<tr>
<th>Quality criteria</th>
<th>Definition</th>
<th>How I addressed this criterion</th>
</tr>
</thead>
</table>
| Credibility      | This relates to the internal validity and consistency of the research and the way it is communicated. Credible research presents well-founded and plausible results [105] | • Triangulation of data collected using different methods  
• Prolonged engagement with participants and agencies (albeit not with the level of immersion I would have liked) via interviews and field observations  
• “Thick descriptions” [76] developed in the case studies  
• Cross-case comparisons to identify and test emergent theories in different contexts  
• Sampling for maximum variation in roles, views and experiences of research use  
• Providing a researcher credibility statement [106] (see section after this table) |
| Transferability  | Transferability relates to external validity or generalisability, the goal being to provide sufficient information for others to determine if and how findings from one study might be applicable in other settings and circumstances | • Using realist approaches to develop tentative propositions at the level of middle range theory which illustrate how intervention strategies functioned in specific contexts  
• Multiple case study comparison which provides a strong basis for theorising due to testing hypotheses about interactions in different contexts. [98] Even though the case studies are not publicly available, the research analyses were contextually grounded and these factors are reported with enough detail that others can see how findings might apply in their settings [20, 25] |
| Dependability    | Dependability can be likened to reliability. Here, the goal is to make the research process transparent and rigorous | • Accountability to the process evaluation contributors  
• Reflexive practice via consideration of myself as the research tool in analytic memos [25]  
• Maintaining an audit trail. In two studies (Chapters Four and Nine), I include additional files that provide a hyperlinked ‘evidence’ link between the tabulated results and the rich data that sits behind them  
• Describing methods in detail, and providing data collection tools and evidence of evolving analyses (Chapter Seven).  
• Statement of “researcher effects” (see below) [116] |
| Confirmability   | This recognises that research can never be objective but that “findings should represent, as far as is (humanly) possible, the situation being researched rather than the beliefs, pet theories, or biases of the researcher” [115:93] | • Comments by participants on draft findings (Chapter Eight)  
• Co-coding and independent analyses, where theoretically appropriate [102] (Chapters Four and Nine)  
• Data analyses workshops and on-going consultation with the process evaluation contributors who challenged my interpretations and obliged me to address differences of views in the final manuscripts  
• Adherence to principles of constant comparison as described earlier |
Researcher credibility

Where the researcher is the primary instrument of data collection and analysis, researcher credibility is a quality consideration [106], and information should be provided that enables others to assess their conduct of the research. [25, 112] Table 2.2 provides an overview.

Table 2.2. Overview of key skills and experience, and implications for this research

<table>
<thead>
<tr>
<th>Key skills and experience</th>
<th>Implications for this research</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am an ex-child protection social worker, counsellor and play therapist</td>
<td>I am experienced in methods that are core to qualitative research such as active listening, dispassionate observation, and reflexive practice⁴. I also know what it is like to be a frontline practitioner</td>
</tr>
<tr>
<td>I studied research methodology and the philosophy of science during my MA, focusing on qualitative research, and then worked as a qualitative researcher</td>
<td>I am qualified to conduct qualitative research in health agencies, but have limited quantitative research skills so I may overlook opportunities to ask quantifiable questions, or to collect quantitative data that could complement my qualitative work</td>
</tr>
<tr>
<td>I have worked in program management and policy development in state and federal health services</td>
<td>I have some personal understanding of the challenges in health policy and program development and implementation</td>
</tr>
<tr>
<td>I have conducted research into how researchers influence, and are used within, public health policy [94, 117, 118]</td>
<td>This work gave me a background in the fields of health policy research and knowledge mobilisation</td>
</tr>
<tr>
<td>I continued to study, present and collaborate on aspects of research theory and practice, with a focus on qualitative research and evaluation, and more recently, realist approaches</td>
<td>This enables me to conduct qualitatively-orientated theory-driven research and evaluation, including using realist approaches</td>
</tr>
<tr>
<td>I have sat on a Human Research Ethics Committee for several years</td>
<td>I have a sound understanding of, and commitment to, ethical research practices</td>
</tr>
</tbody>
</table>

The independence of my research is a further credibility consideration. As a member of the team employed to work on SPIRIT I was not independent, but some measures were in place to aid my critical thinking about the trial. I was not involved in the design or implementation of the intervention or measures. In keeping with the study protocol, was not privy to information about outcomes, and did not provide formative feedback from the process evaluation during the intervention or take part in conversations about implementation improvements. This gave me some distance from the research team and prevented me from investing in any specific aspect of the intervention or measures. Importantly, the process evaluation contributors were involved in all my analyses. They provided different theoretical perspectives, played an active part in developing and critiquing hypotheses, and enabled me to receive expert advice from colleagues who were not involved (or invested) in other parts of the study.

⁴ Reflexive practice includes awareness of personal biases. An example of this is when I noticed my belief that policymakers would want to know how to access research. This assumption became apparent in the first few weeks of data collection when an interviewee made it clear he had no such desire. I then attempted to identify other assumptions (and over-generalisations) in the interview questions and in my views.
Researcher effects

Only two occasions out of 54 observations raised concerns about researcher effects. In one case the participant had not been aware that the workshop was part of a trial and was merely surprised by the evaluation, in the other the participant was suspicious about how evaluation data would be used. It transpired she had participated in a previous study where data was misused. Unfortunately, she could not be reassured, even by offers to cease the evaluation of that workshop, and her continued antipathy disrupted the session.

There was an occasional suggestion that my presence bothered SPIRIT workshop presenters. I attempted to minimise the sense of surveillance by sitting out of eye contact, and being friendly prior to the workshop. I learnt during my social work training while conducting long-term ‘infant observations’ in the homes of new parents that an adult human cannot convincingly act like a fly on the wall, and that friendliness and genuine warmth (i.e. being more human rather than less) is the best way to convince people they can afford to act naturally.

My lack of independence may have had an impact. Participants knew that I was based at the same organisation as the chief investigator, study director and knowledge broker, and they sometimes assumed that I was involved in decisions about intervention design and implementation. In interviews I did not always explain that I had no involvement in the intervention because when I did so it sounded disingenuous—I was a member of the research team, after all—and I suspected this disavowal might be perceived as “the lady doth protest too much”. I was also a researcher asking questions about an intervention designed to increase the use of research, so it is unlikely that I would be perceived as disinterested. Consequently, concern about my views and allegiances may have inhibited full and frank feedback about SPIRIT. Lastly, process evaluation interviews, conversations and feedback forms took people’s time, and this will have added to the overall burden of participation.

Other than the example described above, there was no indication that researcher effects caused policymakers to alter their behaviour in workshops, or that social desirability affected interviews. As in previous interview-based studies with policymakers [94], most participants were generous with their time, keen to ensure they were understood, and matter-of-fact about areas of anticipated disagreement (this is not surprising given that policymakers navigate conflicting views and interests as a core professional skill). The impression was that people who took part in interviews and informal conversations spoke openly5. But of course people with concerns may have not have participated in interviews, or may not have attended the workshops where I was conducting observations and talking with participants.

5 In my experience of this and previous research, policymakers tend to be frank and curious interviewees. But when asking policymakers about current and/or controversial policy endeavours that could place them on a political firing line, they are more likely to act as information gatekeepers and stick to banal ‘on message’ responses. They may also demand such high levels of confidentiality that it prevents meaningful use of the data. See Lancaster K (2016) Confidentiality, anonymity and power relations in elite interviewing: conducting qualitative policy research in a politicised domain. *International Journal of Social Research Methodology*:1-11.
Chapter Three: The challenge of increasing research use in policymaking

“Evidence is the six-stone weakling of the policy world” [in competition with a] “four-hundred-pound brute called politics…”


Why intervene in policy agencies’ use of research?

Over the last few decades there has been an intense global focus on the use of research in policy processes.[4] Scholars draw attention to the disjuncture between the production of research and its role in guiding decision-making, noting the huge waste of knowledge, research effort and public money that results. [120-122] They argue that public policy should be informed by the best available evidence if it is to be both efficient and effective. [123] These arguments have particular heft in health where costs are escalating at an alarming rate due to the triple trends of increased chronic disease, increasingly expensive technologies and longer life expectancy, and where the quality of policies and practices can literally have life and death impacts.

Governments, too, assert the need for policy to be informed by research. The UK’s ‘New Labour’ under Tony Blair and Gordon Brown, followed by the conservative Cameron government [124]; the Obama administration in the US [125]; Justin Trudeau’s government in Canada [126]; and recent Australian governments of different political hues [127] all espouse the broad principle that policy should be informed by robust science when possible, and to some extent have enshrined this in legislation. Governments in low- and middle-income countries, and major international health and aid agencies, are following closely behind. [127-129]

There is no firm evidence base for the claim that greater use of research in policymaking results in better efficiency or effectiveness than decision-making based on informed guesswork, expert hunches or political imperatives. [130, 131] But it does appear that a stronger research culture within healthcare services is associated with benefits to patients, staff and the organisation [30, 132], so it is reasonable to assume that policy agencies may benefit similarly. Unfortunately, our ability to capture accurate information about such outcomes is hampered by the profound difficulty in assessing research impact, including the challenge of defining exactly what we mean by research use within complex policy processes. [28, 133, 134] We are yet to devise robust methods that measure, or otherwise systematically capture, the impact that research has on policies and practice. [135] Such efforts are complicated by the often circuitous and poorly theorised pathways through which research can influence decision-making [90], and by the contextual specificity of each case. [136]

Nevertheless, research demonstrably contributes to a wide range of important social, economic and cultural developments [137] and there are many examples of ‘successful’ policy where research appears to have had a decisive role in improving outcomes. [125, 138] A lot of established policies and the research that informed them are simply taken for granted because they have become ‘obvious’ social goods. Examples of these include building and engineering safety standards, environmental health regulations, motor vehicle design specifications, and smoke-free workplaces [139] And there is general agreement that, “Research
Evidence has undoubtedly been crucial in formulating countless global health policies which have saved many millions of lives.” [4:97]

Consequently, using reliable research judiciously in the development of policy is not a panacea, but it is our best option. As O’Dwyer puts it: “We must acknowledge that ‘evidence-based policy making’ is not synonymous with ‘good policy making’, but that evidence-based policy making is more likely to be good policymaking.” [131] Efforts to improve the use of research in health policymaking, therefore, have the potential to increase the quality of programs and policies, which is likely to increase the efficiency and effectiveness of health services. [123, 130]

Challenges to improving the use of research in policymaking

Interventions to change professional behaviours in complex settings face many challenges. This section focuses on four aspects that may present the biggest challenges for interventions that aim to increase or otherwise improve the use of research in policy agencies, and considers the implications for intervention design. For the sake of completeness, the models of research use that are commonly identified in the literature are described in Appendix 2. This provides a background to the terrain and highlights the limitations of conceptualising research use as a linear rational process that has direct (instrumental) impacts on policy decisions.

1. Human rationality is limited

The peculiarities of human information processing present several problems for instrumental models of policymaking [125, 140] For example, the literature on bounded rationality [141] argues that people’s ability to analyse research objectively and blend it with other forms of evidence is impeded by cognitive and contextual constraints resulting in satisficing: accepting a level of knowledge that seems good enough, rather than optimal. [141, 142] Further, dealing with the complexity of policy problems forces individuals and institutions to develop simplified accounts of the world. To do this, vast amounts of information, particularly any that conflicts with current views, must be omitted. The resulting “social construction of ignorance” makes uncertainty more manageable. [143] Satisficing and socially constructed ignorance mean that potential agendas, outcomes and alternatives are inevitably neglected. [91, 141, 143]
Policy rationality is not limited only by cognitive constraints; multiple factors impair decision-making, such as:

1. Information characteristics such as problem definition, research quality, availability, author credibility, accessibility, policy relevance and applicability, choice overload, and the inability of much research to reduce uncertainty about what action is best [125, 140, 144, 145]

2. Information mobilisation characteristics such as timeliness, dissemination strategies, source credibility, knowledge brokerage, framing, and content format [28, 88, 146, 147]

3. Individual policymaker characteristics including ideology, judgement heuristics, emotional response and investment, worldview, current work role, and personal and professional experience (including research skills). This means that research use may depend more on individual capacities and interpersonal relationships than on the attributes of the research. [79, 148-150] As Stone puts it, “In our various social and political roles, we act largely according to prior attitudes and beliefs rather than new information”. [151] Thus, even with the best evidence, policymaking cannot be “debiased”. [140]

4. Organisational characteristics such as constrained time and resources [91], the agency’s culture of evidence use, its usual level of risk-taking, knowledge management structures, and rules and procedures. [38, 144, 147, 152] When these are not supportive of research there is little chance that individuals will engage in rigorous use of evidence. [28] Rather, staff are inclined to obtain information from trusted internal sources because this is more likely to support organisational interests. [144]

5. Policy issue characteristics include the specific policy arena and goals [79, 144], how the issue has been “problematised” [153], high issue salience, the availability of plausible alternative policy options, the availability of alternatives that are more ‘extreme’ or ‘attractive’ (thereby allowing policymakers to default uncontroversially to a ‘middle way’), and the inherent uncertainty in tackling complex social problems. [125, 140]

An important body of work draws attention to the social and processual nature of knowledge as something that evolves through experience and via negotiation with peers, and is in continual flux [31, 74-77, 82, 154]; thus professional knowledge has a “social life” in which it intermingles with many forms of information, being re-interpreted and re-embodied as it evolves. [31, 76, 155] If meaning materialises through communication [154] our conceptualisation of rational decision-making must incorporate far greater plurality and ambiguity than it does at present. This includes recognising that membership of workplace communities (epistemic cultures) shapes professional identity and the kinds of knowledge that are valued. It also raises questions about our continued focus on “explicit, canonical knowledge” [31], and how we can best integrate research findings with existing beliefs and practices. [31, 74-77]

Given these arguments, rather than criticise policymakers for failing to use research, we may make more progress by recognising the contingent rationality in their behaviour and attempting to work with this. In other words, “Instead of being shocked that policymakers sometimes ignore evidence, we need to better understand the values, beliefs and processes that guide their decisions”. [4:98] One way that research utilisation interventions could do this is by conducting careful situational analyses and/or taking a participatory approach to program development and implementation.
2. Policymaking is complex

Policymaking is a messy, conditional, serendipitous and profoundly context-dependent process \[28, 35, 62\] in which decision-making is highly unpredictable—it may be weighed down indefinitely or "everything happens at the same time". \[90, 93\] This reality contrasts strikingly with procedural or staged models of decision-making that depict distinct tasks in a ‘policy cycle’. Policy is no longer thought of as a boundaried event but as an entanglement of fuzzy processes \[28\], and the way these processes unfold in relation to context will affect what research is considered, and if and how it is used. \[156\]

Policy is often trying to tackle wicked problems, i.e. intractable social ills with multiple causes that are not ‘owned’ by any single jurisdictions, which are characterised by conflicting values and views, and where the boundaries of the problem and its solution are uncertain. \[57, 71, 91\] Examples include obesity, climate change and Indigenous disadvantage. \[157\] Research can seldom provide a solution because the problems evolve and "need to be worked on continually rather than waiting for the right answers to emerge from a period of inquiry by which time policy and practice may have moved on". \[130:586\] Further, "Even when research does point to practical action, it takes more than knowledge and ideas to make policy. It takes imagination and creativity to transform ideas into workable proposal and it takes mobilization and political support to turn the proposals into policy". \[158:289\] This complexity undermines the authority of research and other forms of evidence as instrumental decision-making tools.\(^6\)

Nevertheless, complex problems often demand multifaceted policy responses \[88, 159, 160\], meaning that a single problem may require “… legislation, regulation, enforcement, cross-sectoral cooperation, incentives, attempts at shifting sociocultural norms, and programs and services made up of multiple packaged interventions requiring supportive financial mechanisms, infrastructure, workforce, and governance structures to be integrated into complex, dynamic health and political environments”. \[161\] The relevance and utility of

\(^6\) Realists argue that the uniqueness of wicked problems has been overstated and that there are always lessons that can be applied from one policy or program to another, providing the right methods of inquiry are used.
research to each activity may determine whether the research is even considered and, if so, how it is used. As Nutley et al, explain, “research knowledge cannot simply be adopted by the political system: instead it needs to be adapted, recreated and transformed before it can be used.” [28:100]

Weiss cautions that the complexity of policymaking constrains choice to such an extent that policymakers seldom view themselves as ‘decision-makers’, arguing that their need to accommodate competing interests and to work within a bureaucratic hierarchy prevents any individual from making a unilateral decision. Even at the highest levels policymakers report that their decisions are often “prefabricated” in that they are simply rubber-stamping recommendations that have arisen through the tiers below them. [162:268] Given that the way in which policymakers conceptualise their work will affect what kinds of information they seek and use [162], and that knowledge mobilisation efforts in policy are often framed in terms of decision-making, this area requires better understanding. [e.g. 35] Unfortunately, the research utilisation literature still offers surprisingly little insight into how policy is actually made, implemented and evaluated [90], possibly because researchers—the primary authors—often have poor understanding of policymaking themselves. [163, 164]

Arguments about the lack of efficiency and agility in policymaking are often attributed to the legendarily cumbersome decision-making architecture of bureaucracies, which are portrayed as Kafkaesque behemoths populated with paper-pushing pedants who repress individualism and innovation. [7] But this may be as much to do with complexity as it is with managerialism. Policy agencies are multifaceted, heterogeneous, nested within larger systems of government, subject to multiple constraints including intense pressure from diverse stakeholders and a frequent need for rapid yet uncontroversial response. [56, 165] They are also tasked with effecting change in highly complex open systems. In health, these include hospital and community services, and entire health systems, which themselves intersect with other systems such as education, law, transport, housing, and social welfare. [29, 166–169]. In short, policy work is “...embedded in intricate networks of physical, biological, ecological, technical, economic, social, political, and other relationships” [166:505].

Attempts to improve how research is used must take account of this complexity [57–59, 170, 171]. Hawe [69] argues that the trick of an intervention is to ‘couple’ with this shifting context and adapt reflexively to influence change in the desired direction. This requires collaboration with participants, fostering networks, and evaluation techniques that can take account of greater flexibility [170]. For example, fidelity assessment needs to focus on intervention function rather than form (this is explored in Chapter Seven) [18], and developmental evaluation can be used to identify emergent effects so that intervention strategies, or their implementation, can be adjusted to minimise negative change trajectories and maximise positive ones [172].

---

[7] Popular representations of bureaucracy tend to be caricatured and harsh. For example, the 19th Century philosopher John Stuart Mill observed that, “The disease which inflicts bureaucracy, and what they usually die from, is routine” and more recently the economist Thomas Sowell said, “You will never understand bureaucracies until you understand that for bureaucrats procedure is everything and outcomes are nothing.” Many novelists hold similar views. In Heretics of Dune (1984) Frank Herbert writes, “Bureaucracy destroys initiative. There is little that bureaucrats hate more than innovation, especially innovation that produces better results than the old routines.” While in The Shadow-Line (1917) Joseph Conrad observed, “The atmosphere of officialdom would kill anything that breathes the air of human endeavour, would extinguish hope and fear alike in the supremacy of paper and ink”. The terms bureaucrat and bureaucracy are often used pejoratively, for example, see https://en.wikipedia.org/wiki/Bureaucracy and www.govloop.com/is-bureaucrat-a-bad-word/.
3. Policy is, and will always be, political

"Here are some of my policy assumptions. Find something to base them on."

© Bernard Schoenbaum/Condé Nast

According to Greenhalgh and Russell, terms such as “knowledge translation” and “getting evidence into policy” are seductive but inaccurate metaphors because policy is not informed by objective evidence about problems that are “out there” waiting for solutions. [61] Rather, it constructs problems and solutions through a process of argumentation which is “characterised by bargaining, entrenched commitments, and the interplay of diverse stakeholder values and interests”. [173] As Head argues, “The policy-making process in democratic countries uses the rhetoric of rational problem-solving and managerial effectiveness, but the policy process itself is fuzzy, political and conflictual”. [173:83]

In this process, decisions about what evidence to use and how to use it focus not on probabilities but on plausibility, persuasion, utility and appeal [34, 89, 174] or, as Gibson puts it, “framing and taming” policy problems. [in 61:315, 62] This shifts us from “equating rationality with scientific, technical procedures to considering rationality as a situated, contingent human construction” where the process of deliberation itself is critical [174:63]. Echoes of this can be found in the assertion by Weiss [158] and, more recently, Smith [92] that ideas rather than evidence are often most potent in policy processes.

There is also a tension between evidence and values, i.e. between what we know versus what we hold to be important. [175] Values are at the core of policymaking because it “...is largely about deciding who gets what and who pays [and how] ...to calculate ‘trade offs’ between conflicting demands and priorities”. [176] Democratic values such as freedom, equity, choice, security and justice form the basis of this calculation but are subject to contradictory interpretations [151] yet, in the end, “...facts and values are so intertwined in policy-making that factual arguments unaided by persuasion seldom play a significant role”. [36:8] Within such a context blending research with other forms of information to reframe problems, persuade stakeholders and dominate arguments seems entirely rational. [28, 88, 94]

These points suggest that research utilisation interventions in policy agencies need to recognise that research can make a valuable contribution, but “…other legitimate influences on policy (social, electoral, ethical, cultural, and economic) must be accommodated”. [177:277] Indeed, some researchers argue that the
attempts to repress these other influences in favour of research are not only naïve, but “dystopian because the assumption is that this multiple stakeholdership, this pluralism, is somehow less desirable than a situation that privileges evidence and the producers and purveyors of evidence, i.e. us!” [60]

Since policy cannot be depoliticised, perhaps researchers can find ways to accommodate political processes more effectively? This might include providing findings in the form of compelling narratives that frame them in relation to current concerns [33, 178, 179]; tactical collaboration with policymakers via advocacy groups and in media campaigns; and learning about policy realities that include the inevitable need for compromise as values, interests, feasibility concerns and other factors are incorporated into the evolving policy process. [118] In short, researchers—including research utilisation intervention designers—must find a way to avoid the contradiction often observed in the literature of admitting that policymaking is political, yet proceeding as if it can somehow be transformed in to “an apolitical rational decision-making process”. [34, 89, 124]

4. Evidence is contested

Given the arguments above, it will come as no surprise to read that evidence is contested. However, in order to understand the research findings that are presented in later chapters, it may be helpful to consider “What counts as evidence?” [180] or perhaps, to be more precise, what counts as evidence for whom and in what
circumstances? [181] Many of the main points can be found in, or inferred from, the earlier sections of this chapter. These include: evidence comprises much more than research, e.g. stakeholder perspectives, public opinion, values and contextual considerations; policy cannot be based on research alone; knowledge, including scientific knowledge, is socially (and politically) constructed [182], thus research is not simply "static information that can be adopted or rejected" [124], rather it has to be reformulated and given legitimacy within communities of practice where tacit experiential and context-specific knowledge can be incorporated. [31, 183] Two further, related, points should be noted:

**Research must be fit for purpose**

There is often a poor fit between available research and the needs of policymakers, particularly in health. [130] This is partly because current research paradigms struggle to provide policy relevant answers. [184] All parties acknowledge that robust research conducted according to appropriate scientific principles is a good thing, but there is considerable disagreement as to what is appropriate; there is no universal gold standard methodology for policy-useful research.

However, scholars do make suggestions for conducting research that would be more fit for purpose. They call for less descriptive, behavioural and risk factor modification research—we are already "awash" with it [130]—and more social, environmental and economic interventions that can address key knowledge gaps. [130] Many argue for fewer randomised controlled trials that have limited value for delivering and understanding how programs work in natural settings where system level factors are critical. [18, 66, 184-187] Instead, there should be greater emphasis on 'real-life' and 'real-time' applied research [130] using pragmatic designs. These are thought to have greater capacity to effect change as they are more likely to offer generalisable results. [184, 186] Given that we tend to overestimate how well research can 'travel'—and thus be applied—from one context to another [188], many argue that there should be greater focus on reporting implementation processes and contexts (including challenges) so that others can determine what support factors were operating and compare it to local conditions. [25, 73, 186, 189] Process tracking and evaluation that focuses on causality is strongly recommended. [15, 188] Throughout these processes, researchers need to engage with policymakers, practitioners and other stakeholders with a view to partnerships and co-production. [32, 156, 190-193] Lastly, we are advised to package research accessibly, tailoring it for specific audiences and emphasising "actionable messages" rather than methodology. [190, 194] Creative dissemination, including engagement with the media and advocacy groups, can help get these messages beyond the confines of academia. [88, 90, 130, 179]

This suggests that research utilisation intervention trials in policy agencies may benefit from targeting local research generation (funding, commissioning and in-house development) and partnership approaches across policy, research, the not-for-profit sector and service organisations as well as the more common focus on accessing and applying research that is already ‘out there’. Intervention approaches that might serve these goals include action research, rapid cycle interventions and other quality improvement strategies, and soft systems methodologies, all of which focus on local need in context, and use participatory methods to identify problems and solutions.

**The truth is up for grabs**

Much of the discourse on research-informed policymaking is concerned with fundamental conflicts about the nature of evidence and its relationship with policy practices. [28, 195] In her criticism of the "rationality project", Stone [151] [125] dismisses the distinction between information, evidence and propaganda, arguing that what we think of as facts, including research findings, are produced through social processes involving contested choices. This may partially explain why vast bodies of evidence can have little impact on policy or practice. [196] In policymaking, the status of any form of knowledge, including research findings, is
uncertain [133], “Policy making remains a contested and contingent site where various types of evidence and forms of knowledge come into play and come up against power relations and established hegemonies about what constitutes the ‘truth’. [in 131:130] Consequently, facts alone, no matter how rigorous or cumulative, are seldom sufficiently persuasive to change minds. [197, 198] This is not to claim that all knowledge is relative, but that it is rubbery, and it is interpreted differently according to beliefs and interests. Indeed, the terms used to describe truth—e.g. objective, real, unbiased—are themselves unstable and subject to quite different interpretations. [47] Perhaps we should not be surprised that the term post-truth is now in common usage.\(^8\)

When it comes to research, even if reliable quantitative data are seen as factual, what those data mean in the context of policy decision-making is a matter of opinion. Some question whether such data should be called factual given the lack of genuine objectivity in how scientific questions are formulated, which data are collected and how they are analysed. [107, 182] Despite general agreement that science produces “conditional truths” at best [160], scientists themselves have very different ideas about how truth-like different forms of research findings can be. For example, the so-called evidence hierarchy [199] is criticised by many [34, 200, 201], and can lead to an “inverse evidence law” where the forms of research most valued by researchers are least likely to be used by policymakers. [202] Others claim that most published research is false because: it lacks sufficient statistical power; and has biases in design, data, analysis [203]; or argue that the quest for neutral evidence is a fool’s errand as we will always observe, describe, theorise, and explain the world through our own interpretive lenses. [204] This is supported by empirical studies that find the same research can be perceived and used quite differently by different policy groups [205] and, indeed, by different reviewers. [206]\(^9\)

These points suggest that research utilisation interventions need to wrestle with questions about the forms of research they consider legitimate in terms of methodology, source, funding and political acceptability, finding a middle-way between policy-utility and research rigour. We need to ask questions such as ‘How can policy questions be addressed meaningfully and robustly?’; ‘What forms of synthesis are both reliable and relevant?’ [195] and, if the intervention seeks to build capacity, ‘What skills and outcome expectations are most appropriate in this context?’

---

\(^8\) Post-truth was the Oxford Dictionary’s word of the year, 2016. It is defined as “relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief”. The dictionary editors said that use of the term post-truth had increased by up to 2000% in 2016 compared to 2015, and attributed this spike to the UK’s EU referendum and the US presidential election. (The Guardian, 16 Nov 2016, available at: https://www.theguardian.com/books/2016/nov/15/post-truth-named-word-of-the-year-by-oxford-dictionaries). See also http://wapo.st/2fVd33w?tid=ss_tw which describes post-truth in the US election and introduces the useful term “truthiness”: the phenomenon of “believing something that feels true, even if it isn’t supported by fact.”

\(^9\) This does not mean there are no truths. We need a balance between recognising social constructions and “veriphobia”—fear of the truth—which condemns all attempts to pursue truths as sinister or naïve (Oakley et al., 2005).
The need for research-informed policy

We have seen that research may not be able to reduce uncertainty by providing definitive answers, politically neutral facts or unproblematic policy actions [125, 144], but there are many things that it can do. Research can help:

- **Define problems and (re)frame them** [152], acting as “a powerful vocabulary that shapes the definition of issues” [162:276], and draws attention to overlooked possibilities. Given that the way a problem is defined guides the consideration of solutions [153, 207] this is a critical contribution.

- **Develop solutions and clarify objectives.** For instance, research can clearly delineate costs and benefits, enabling informed transparent debate about trade-offs, and establishing criteria on which policy performance should be assessed. [127] This can also help policymakers avoid “solutioneering”—jumping to solutions without adequately identifying the problem and what causal factors need to be considered. [127]

- **Articulate options** by showing what (and how) strategies have worked elsewhere and which option may be most appropriate in local circumstances. [152]

- **Test hypotheses for informed decision-making.** [125] This includes piloting programs to better understand implementation-context dynamics [7], and the increasingly sophisticated use of modelling that can present likely outcomes for policy decisions and potential counterfactuals that help to allocate policy resources for maximum gain. [127, 161]

- **Maximise learning from current policy and program efforts.** Rigorous evaluation can reveal what worked and what did not work (and why), avoiding costly mistakes in re-investment and scaling up [127], and clarifying what is meant by policy ‘success’ or ‘failure’. [208] Crucially, this includes showing how policies affect different groups of people, helping to target initiatives more accurately and informing equity-based decisions about improvements or alternative strategies [152, 209], including adapting programs for different contexts. See, for example, the increasing sophistication and well-established impact of research-informed mass media campaigns on public behaviours relating to the use of tobacco, alcohol and other drugs; heart disease risk factors; sexual health; road safety; cancer prevention; child survival; and organ and blood donation. [210] Research can also show where ‘common sense’ policies have unintended or perverse outcomes such as recidivism programs that encourage reoffending, and conservation policies that promote land clearance. [127]

- **Highlight opportunities** to adjust policies or use new approaches based on data about changing trends, incidents or crises. [79, 173]

- **Critique current, pending or potential policy directions** [130, 152] which can provide leverage for external criticism and demands for increased accountability. [127]

Research-informed policy, then, is a fragile but worthy objective. As Hunter puts it, “While public health is both an art and a science, it should not be an act of faith.... Evidence-informed public health therefore seems to be a reasonable, modest and achievable goal”. [130:586]

Getting research into policy

How should we advance this goal of research-informed policy? Many policymakers feel they have limited skills in accessing, appraising and using research [28], but there is scant evidence about what capabilities actually exist, which are most needed, or what strategies might be most effective in building capacity to use research. [191] Some policymakers report that they struggle to identify research and evaluation (including
work that their organisation has funded, commissioned or conducted internally), so organisational knowledge management and other support systems may be suboptimal. [133] There is also scope to better apply this research once it is in hand. [130]

However, we cannot assume that policymakers are enthusiastic about the prospect of increased capacity to use research in their work. In fact, research may be spurned by those we seek to influence in both policy and practice. For example, Nelson et al. [211] found that some tertiary educated mental health professionals not only lacked interest in using research but were actively hostile to the idea. Others note “pockets of antipathy” to research within service organisations. [28] Some policy studies have found low levels of interest in or understanding of research by varying proportions of policymakers [28, 162, 212], and a worrying level of generalised distrust in researchers and their work. [79, 147, 213, 214]

There are clearly considerable challenges ahead and we have to accept that policies will seldom simply operationalise research findings. However, there are many opportunities for more modest improvements. [130] The strategies we use to support the use of research in policymaking may end up being “… more of an art than a science, requiring considerable amounts of judgement and luck” [88:29], but it is an art that we must try to master.
Chapter Four: A review of research utilisation interventions in policy agencies

“[We must] get to grips with the theory underpinning EIPM [evidence-informed policy making], especially given criticism that EIPM research and interventions fail to take account of the messy reality of policy processes. Building theoretical insights into interventions can help... avoid common traps and design programmes that are more likely to lead to change.”

Punton et al., 2016 [152:4]

Overview

One of the most consistent findings from health services research is the failure of much research translation to effect desired changes in practice and policy. [194] Chapter Three illustrates many of the reasons for this in the policy sphere. But what does—or could—positively affect the use of research in policymaking? And in what circumstances?

The following realist scoping review maps intervention activities 2001-2016 that have been used to increase the capacity of individuals and organisations to access and use research for policymaking. Although the evidence is slim, the review posits some possible causal mechanisms, and explores potential implications for other interventions. Theoretically-informed understanding of interactions in these interventions—even when it is this tentative—can offer clues about how and why SPIRIT functioned as it did across the six sites, and about how future interventions are likely to function in various circumstances.
Manuscript

Haynes, A, Rowbotham, S, Brennan, S, Williamson, A, Moore, G, Redman, S, & Butow, P (accepted for publication at Health Research Policy and Systems) What can we learn from interventions that aim to increase policymakers’ capacity to use research? A realist scoping review.

Abstract

Background
Health policymaking can benefit from more effective use of research. In many policy settings there is scope to increase capacity for using research individually and organisationally, but little is known about what strategies work best in which circumstances. This review addresses the question: What causal mechanisms can best explain the observed outcomes of interventions that aim to increase policymakers’ capacity to use research in their work?

Methods
Articles were identified from three available reviews, and two databases (PAIS and WoS), 1999-2016. Using a realist approach, articles were reviewed for information about contexts, outcomes (including process effects), and possible causal mechanisms. Strategy+Context+Mechanism=Outcomes (SCMO) configurations were developed, drawing on theory and findings from other studies to develop tentative hypotheses that might be applicable across a range of intervention sites.

Results
We found 22 studies that spanned 18 countries. There were two dominant design strategies (needs-based tailoring and multi-component design), and 18 intervention strategies targeting four domains of capacity: access to research, skills improvement, systems improvement, and interaction. Many potential mechanisms were identified, and some enduring contextual characteristics that all interventions should consider. The evidence was variable, but the SCMO analysis suggests that tailored interactive workshops supported by goal-focused mentoring, and genuine collaboration, seem particularly promising. Systems supports and platforms for cross-sector collaboration are likely to play crucial roles. Gaps in the literature are discussed.

Conclusions
This exploratory review tentatively posits causal mechanisms that might explain how intervention strategies work in different contexts to build capacity for using research in policymaking.
Introduction

There is widespread agreement that the use of research in policymaking could be improved, with potentially enormous social gains. [215, 216] There are disputes about the extent to which research can inform policy decision-making, and about what forms of research knowledge should be most valued and how they should be applied, but these are underpinned by a shared belief that the effective use of relevant and robust research within policy processes is a good thing. [61, 217-219] Specifically, that research-informed policies can help prevent harm, maximise resources, tackle the serious challenges facing contemporary healthcare, and otherwise contribute to improved health outcomes. [220-223] Ensuring health policymakers have the capacity to access, generate and use research in decision-making is a priority.

Despite a rapidly growing body of literature about the use of research in policymaking, we have limited understanding of how best to help policymakers use research in their day-to-day work. Partly, this is because most of the literature is either descriptive or theoretical. Descriptive studies often struggle to identify findings that are transferable to other settings, which can limit their utility for informing intervention design. [224] Theoretical studies have produced many models, concepts and frameworks, but these are often hard to operationalise [134] For example, Field et al. [225] found that one such framework, while frequently cited, was used with varying levels of completeness and seldom guided the actual design, delivery or evaluation of intervention activities. The authors conclude that prospective, primary research is needed to establish the real value of theoretical models and tools. Yet testing specific strategies to increase or otherwise improve the use of research in policy processes is relatively underdeveloped. [226, 227] Consequently, we have a plethora of ideas about what may or may not support research-informed policymaking, but little robust empirical knowledge about what strategies are effective in which contexts.

This paper brings together information about interventions designed to build capacity for using research in policy processes, and explores possible transferable lessons for future interventions.

Using research in policymaking

The concept of research-informed policy has emerged from multiple disciplines with different paradigmatic influences, leading to debates about what we should expect of policymaking and the role of research within it. [38, 96, 218, 228] In summary, many reject what they see as inaccurate assumptions about the extent to which policymaking is a linear technical-rational process in which ‘objective’ academic research can be used instrumentally (i.e. to direct policy decision-making). [61, 228] This argument is premised on policy being a rhetorical arena [34] where, often, facts are uncertain, values are contested, stakes high and decisions urgent. [229] Such views challenge the expectation that improved access to research, or greater capacity to use it, will result in increased use. [230] Indeed, several studies show that individual attributes and contextual factors frequently prevent the use of apparently helpful evidence. [e.g. 149, 231, 232, 233] Some go further, questioning the assumption that there is a policy-research ‘gap’ that needs ‘bridging’, or that more use of research in policymaking will necessarily produce better health outcomes. [35]

Counter arguments highlight the enormous number of policymakers who are actively and effectively engaged (and often qualified) in using research; the success of strategies for improving research use such as policy dialogues, rapid review programs, and partnership approaches [234-240]; and the many cases where research has demonstrably influenced policymaking with positive results. [216, 240-242] From this perspective, research is only one source of evidence amongst many, but it is a critical source that has the potential to guide agendas, maximise the functioning of programs, support governments and service providers to act in the public interest, and hold them accountable when they get it wrong. [216, 219, 243, 244] As Lomas puts it, “…the goal here is not for the imperial forces of research to vanquish all other inputs.
Building capacity to use research in policymaking

Capacity building is conceptualised as a suite of strategies that seek to “increase the self-sustaining ability of people to recognize, analyse and solve their problems by more effectively controlling and using their own and external resources.” [de Graaf 1986 in 245:100] Thus effective capacity building interventions facilitate not merely technical skills development, but increased agency and agility. [246]

Capacity is a multi-dimensional concept spanning different levels—individual, interpersonal, organisational and environmental [152, 247]—each of which is likely to require quite different intervention strategies and evaluation methods. [248] Capacity building in policy agencies uses “a variety of strategies that have to do with increasing the efficiency, effectiveness and responsiveness of government.” [Grindle 1997 in 246:212] In this context performance and accountability are essential. [147, 152, 218] Greater capacity to use research can enhance the former by increasing the effectiveness of public policies and programs, and the latter by providing an independent and scientifically verified rationale for decision-making. [152, 249, 250] In this study, we use the term ‘research’ broadly to include collections or analyses of data, or theory, found in peer reviewed papers, technical monographs or books, or in grey literature such as internal evaluations and reports on authoritative websites; or presentations or advice from researchers.

Multiple dimensions of capacity affect the use of research in policymaking. At the most concrete level policymakers must to be able to get hold of useful research. This is research that: (a) addresses policy problems, including costs and benefits, and produces findings that can be applied to local decision-making; (b) is accessible, i.e. readable, timely and easy to get hold of (not buried in information tsunamis, or located behind firewalls); and (c) which has policy credibility, i.e. is conducted with sufficient scientific rigour to render it reliable, but is also methodologically fit for purpose and communicates the findings persuasively. [145, 251, 252] So there is a scope to enhance the conduct, presentation and synthesis of research itself, as well as the means by which policymakers access it. [253]

Policymakers may also need specialist knowledge and skills to access, appraise, generate and apply research in their work. Although many have substantial skills and experience in these areas, others do not [254]; they lack confidence and want training. [250, 255] Individuals’ beliefs about the value of research and requirements of different policy roles are also considered to be important mediators of use. [80, 253]

Organisational capacity can constrain or enhance research use, irrespective of individual capabilities. [256] Institutional infrastructure, resourcing and systems, leadership and the underlying workplace culture, have an enormous impact on practice norms and expectations, and opportunities for skills development and application. [13, 21, 253, 256, 257] Organisational culture is notoriously hard to access and transform [258] but is considered to be a fundamental indicator, and facilitator, of research-informed practice. [152, 259] This meso level is, in turn, impacted by the wider institutional systems in which policy organisations operate. [254] For instance, dominant views about what forms of evidence are valued and how they are sanctioned or incentivised in policy processes shape what capacities are needed and how they operate, but this sphere remains largely outside the scope of intervention trials. [152, 247]

The quality of the relationships between policymakers and researchers is also (and increasingly) seen as critical for improving the development and uptake of useful research for policymaking. [214, 239, 260, 261] Here, capacity building focuses on forging or enhancing connections across a spectrum of interactivity from information exchange forums to the use of knowledge brokers through to formal partnerships and the co-
production of research. [262-264] Individual, organisational and institutional capacity have crucial roles to play in forming and sustaining interpersonal networks. [152, 253]

These dimensions of capacity indicate the breadth and complexity of the capabilities that interventions can address, concerned as they are with products, resources, skills, beliefs, values, systems, institutional structures, boundaries and relationships, most of which have interdependencies.

Aims

This review explores a range of interventions designed to build capacity for research use in policy processes. Outcomes of interest are those related to capacity to use research, including capacity to access and apply research, to work productively with researchers and intermediaries such as knowledge brokers, the establishment of workforce and infrastructure supports, intention to use research and actual use.

Our purpose is twofold: First, to describe the main characteristics of the interventions—the study designs, intervention goals and strategies, implementation settings, participants and outcomes. Second, to consider how process effects and outcomes were generated in those settings (see next section for definitions of these terms) drawing on theory from these and other studies to develop tentative hypotheses that might be applicable across varied intervention sites. Understanding context and theory is essential for understanding how interventions function in general [18, 20, 265], and how research is mobilised. [3, 266, 267] Our aim is to provide a first step towards developing practical and theoretically grounded guidance for future interventions of this type. The research question addressed is: What causal mechanisms can best explain the observed outcomes of interventions that aim to increase policymakers’ capacity to use research in their work?  

Our approach: a realist scoping review

This a realist scoping review. In general, scoping reviews are “… a form of knowledge synthesis that addresses an exploratory research question aimed at mapping key concepts, types of evidence, and gaps in research related to a defined area or field by systematically searching, selecting, and synthesizing existing knowledge”. [268:129-4] In this case, the research question and synthesis were strongly informed by realist philosophy and realist review methods; however, it does not fully adhere to the current criteria for conducting a realist (or rapid realist) review – hence the hybrid term (see Appendix 3 for comparison of scoping reviews, realist reviews, and our methodology).

A realist approach is used because we aim to produce findings with potentially transferable implications for the design, implementation and evaluation of other interventions in this field. Realist reviews attempt to identify patterns that are articulated at the level of ‘middle range’ or program theory. This is thought to be most useful for our purposes because it is specific enough to generate propositions that can be tested, but general enough to apply across different interventions and settings. [15, 53, 269] Realist approaches are methodologically flexible, enabling reviews of broad scope suited to addressing the complexity of interventions in organisational systems [270], and for addressing questions about how interventions work (or not) for different people in different circumstances. [119] This inclusive approach enabled us to capture studies that used innovative and opportunistic strategies for enhancing research use in policy, and diverse methods for evaluating these strategies.

10 A note about our terminology: by intervention we mean a purposeful attempt to bring about some identified change, this may involve the use of one or more strategies. We use the term theory broadly to encompass a range of formally investigated and informal hypotheses about how intervention strategies bring about change, or why they don’t.
Mechanisms, process effects and outcomes

Realist evaluations and syntheses develop and test hypothesised relationships between intervention strategies, implementation contexts, causal mechanisms and observed outcomes. However, contexts, mechanisms and outcomes are not fixed entities in a tidy causal chain, but can shift position depending on the focus of the inquiry; i.e. they function as a context, mechanism or outcome at different levels of analysis and in different parts of a program. [238] For example, if the outcome of interest is research utilisation, then increased capacity is likely to be an important mechanism; but if the inquiry takes capacity itself as the outcome of interest (as we do in this review), the focus will be on the more granular mechanisms that built this capacity. So, in this review we are looking for mechanisms that cause, and thus precede, capacity development. Different foci, and the corresponding shift in where we look for elements of causality, have implications for the review and synthesis of intervention studies, as we now explain.

Where interventions are new, or newly adapted, causal relationships are often examined at a relatively detailed level of granularity that emphasises specific points in the causal pathway. Broadly, we see process evaluation, formative evaluation and much of the qualitative research that is conducted in trials as trying to identify and explain how intervention strategies bring about process effects. These are the range of immediate responses (ways of interacting with and ‘translating’ the intervention) that shape distal responses to the intervention (final outcomes).

Intervention designers have hypotheses about what process effects are needed to achieve desired outcomes. Often these are not articulated, sometimes because they are obvious (e.g. if no one attends a workshop then it clearly cannot be successful), or because these interactions are subsumed in blanket terms like engagement and participation which mask crucial details about how people engaged and participated, or why they did not. For example, intervention studies often report on the importance of champions, i.e. members of an organisation who actively promote an intervention or practice change. When looking at an intervention study as a whole, championing may function as a causal mechanism in that it helps to bring about change, but from a granular perspective, championing can be conceptualised as a process effect because: (a) effective championing mediates intervention outcomes (the influence of champions on organisational change initiatives is well documented), and (b) it is generated by interactions between the intervention, participants and context (i.e. it is caused - people make conscious judgements about acting as champions). Consequently, in order to inform implementation improvements, and program improvement and adaptation, it may be useful to understand the causes of championing in more detail, for example, by asking ‘In this intervention and context, what perceptions and considerations influenced who became a champion and who did not?’. At this level of analysis, process effects are treated as proximal outcomes. We explore this in more detail elsewhere. [271]

Figure 4.1. depicts how these two levels of focus might be applied to a research utilisation capacity building intervention. 4.1.a. illustrates a granular approach where the evaluation focuses on the relationship between immediate perceptions and experiences of the intervention (which function as mechanisms in this scenario) and how they lead to process effects (capacity related responses such as participation in skills development programs, relationship development with researchers, or managers funding access to research databases and other workplace supports). This contrasts with figure 4.1.b. which depicts an evaluation that is more focused on distal outcomes and so takes a higher-level perspective, collapsing the causal detail and blurring the distinction between process effects and mechanisms. From this perspective, many process effects are mechanisms.
Figure 4.1. Different levels of focus depending on the evaluation purpose and outcomes in studies of research utilisation capacity building interventions

Legend: The black dotted lines reflect the focus of enquiry. In 1a, the focus is on immediate responses to the intervention: process effects and the mechanisms through which these are brought about. In 1b, where the focus of inquiry is on more distal outcomes using a higher level of analysis to investigate causality, mechanisms and process effects are functionally the same thing, i.e. proximal responses to the intervention.

In practice, this distinction between granular and high-level foci is usually a question of emphasis rather than a demarcation, and it is not always clear where the focus of an evaluation lies. Although realist findings are often tabulated—which can imply that phenomena exist in strict compartments—the propositions that describe causal pathways tend to suggest greater fluidity and often incorporate process effects and the mechanisms that generate them.

We highlight this distinction here because our findings depend on both the models above. This reflects the different evaluative emphases of the reviewed studies, and their diverse outcome measures. Consequently, we present findings that include intervention strategies, contexts, mechanisms, process effects and outcomes.
Methods

Focus of the review: what and who?

This review is founded on the realist assumption that the success or failure of an intervention depends on the interactions between the intervention strategies and the implementation context. Identifying patterns in these interactions enables us to develop tentative theories about what may plausibly occur in other interventions. [15, 23] Thus, in reviewing the literature, close attention must be paid to the characteristics of the intervention, the people it targets and the settings in which it takes place. To this end, we differentiate between two types of intervention that are often conflated in the literature. This review focuses on research utilisation interventions that aim to increase the effectiveness with which professionals engage with research and use it in their decision-making. We see these interventions as having very different goals to research translation interventions that attempt to modify clinical practice in line with research-informed standards or guidelines. The former will generally attempt to build capacity in some form (e.g. by providing resources, training, reflective forums or partnership opportunities designed to enhance professional practice) and may increase individuals’ agency and critical engagement with research; whereas the latter more often seek to institutionalise adherence to a specific practice protocol which may constrain autonomy and critical reflection. Harrison and McDonald [272] make a similar distinction between the “critical appraisal model” of research-informed practice where participants are encouraged to critique and incorporate concepts in their practice and the “scientific-bureaucratic model” that attempts to regulate practice. The contrasting ‘politics of knowledge’ inherent in these two models are likely to trigger quite different responses in relation to professional identity, self-determination and organisational accountability. [31, 272]

Second, we differentiate between research utilisation capacity building interventions targeting policymakers in government agencies and those targeting health practitioners based in service organisations. While both tackle complex systems with distributed decision-making, we believe that the contextual characteristics of bureaucracies differ from those of clinical practice in ways that may affect the functioning of interventions. This is especially pertinent for interventions that attempt to build research utilisation capacity because the forms of research that are most useful in these contexts are likely to differ. For example, biomedical and clinical research (including randomised controlled trials) may be more useful in healthcare settings, whereas health policymakers might be better informed by research that incorporates scaled analyses of effectiveness, reach and costs. [251] So studies included in this review are limited to research utilisation capacity building interventions that target policymakers.

Search strategy

The literature encompassed in this review was identified in three ways:

1. From three existing reviews of published peer-reviewed papers that reported on the evaluation of strategies aimed at increasing the use of research by decision-makers in public policy and program development. The first two reviews are complementary; one [273] captured papers published between 1999-2009 while the other [274] captured papers published between 2009-2015. These reviews focus on identifying the strategies employed to increase research use and the “factors associated with these strategies that are likely to influence the use of research”. [274] The third review [152], conducted for the UK’s Building Capacity to Use Research Evidence program, focused on primary intervention studies aimed at developing capacity for research use in public sector decision-making. It included studies published between 2003-2014. These sources were deemed to have identified most relevant peer-reviewed papers pertaining to the testing of research utilisation interventions in policy agencies between 2001-2015.

2. From two searches on academic databases: one on PAIS and the other on Web of Science, 2001-2016. See Appendix 3 for the search syntax and filters used, and the rational for selecting these databases.
3. Iterative searches on Google and Google Scholar using keywords and snowballing from citations in previously found papers, reports and journal articles. These were exploratory rather than exhaustive searches, intended to identify papers, reports and commentary that would increase our awareness of diverse perspectives that, in turn, could help us draw theoretically grounded lessons from the intervention studies identified in steps 1 and 2 above. Figure 4.2. depicts the search and exclusion strategy used.
Figure 4.2. Review search strategy

Total of 507 articles retrieved and 486 articles excluded

- Moore et al. 2011 systematic review: 5 articles
  - Included: 3, Excluded: 2
- Campbell et al. 2016 systematic review: 14 articles
  - Included: 6, Excluded: 8
- Punton et al. 2016 review: 18 articles
  - Included: 6, Excluded: 12
- PAIS database search 2001-2016: 215 articles retrieved
  - Included: 3, Excluded: 212
- Web of Science database search 2001-2016: 255 articles retrieved
  - Included: 3, Excluded: 252

Inclusion/exclusion criteria applied

Selected excluded articles from all five sources

21 articles included

Snowball citation searches

1 article identified that had not already been found

22 total articles included

Purposive searches for articles that might shed light on our findings

Selective use of other retrieved articles to frame the review and develop theory
Inclusion criteria

Studies were included provided they met the following criteria:

Interventions. The study reported on an intervention designed to improve or increase policymakers’ capacity to use research in their work. We took a broad view of capacity building that that included strategies for supporting research use as well as strategies for advancing it, so studies were included if they were designed to enhance: access to research; skills in appraising research, generating research (commissioning or conducting it), or applying research; organisational systems for developing and supporting research use; and/or connections with researchers (including partnership work and co-production). Any strategy that employed one or more capacity building activities aimed at individual-level or organisational structures and systems was eligible, irrespective of whether the strategy was part of a research study or initiated/implemented by policy agencies, or a combination. Studies were excluded if they: focused on policy development capabilities in which the use of research was a minor component (e.g. [275]); evaluated one-off attempts to get policymakers to use research for a specific policy initiative; or focused on specific fields other than health, such as education or climate change (e.g. [276]). Studies that addressed the use of research by policymakers in general were included (e.g. [247]).

Populations. The intervention targeted policymakers, by which we mean either (a) non-elected civil servants working in some aspect of policy or program development, funding, implementation or evaluation within government agencies such as Departments of Health and/or (b) senior health services decision-makers, e.g. executives in regional or federal health authorities who have responsibility for large scale service planning and delivery, and/or (c) elected government ministers. We included studies where participants included policymakers and other groups (e.g. frontline clinicians, NGO officers or researchers) and where some disaggregated data for the different groups was reported, but not those in which intervention effects/outcomes were reported in aggregated form, for instance, health staff at all levels took part in the intervention but the data for senior health services decision-makers was not reported separately from that of frontline staff (e.g. [233, 277-283]).

Study design. The intervention was evaluated. This includes process evaluations and reports of ‘soft’ proximal outcomes such as satisfaction and awareness, which we conceptualise in our analysis as process effects. As mentioned, opportunistic evaluations of initiatives that were planned outside the auspices of a research study were included, but studies were excluded if they described achievements or ‘lessons learnt’ but did not explain how such information was captured as part of an evaluation strategy (e.g. [239, 276, 284-286]).

Publication. The evaluation results were published in English between 1999-2016. This date range was judged by the authors as likely to encompass the vast majority of relevant publications in this field, and included the earliest relevant study of which we were aware. [287]

Settings. We included studies from all countries, including low- and middle-income countries. These settings are likely to differ considerably from those of high-income countries (e.g. less developed infrastructure, fewer resources, different health priorities and a poorer local evidence-base [288]) but, together, the studies provide insights into creative interventions and produce findings that may have implications across the country income divide in both directions.

Quality. Studies were excluded if they were “fatally flawed” according to the appraisal criteria used in critical interpretive synthesis. [289] Following Dixon-Woods et al., we used a low threshold for quality pre-analysis due to the diversity of methodologies in our final sample, and because our method of synthesising data would include judgements about the credibility and contribution of studies. [289]
**Analysis**

We took an inductive approach to analysis (e.g. [249]) guided by realist thinking rather than starting with an *a priori* framework or conceptual categories. This was due to the breadth of strategies we were investigating and their diverse theoretical implications. As befits an exploratory realist review, our aim was to identify "initial rough theories" that might explain how and why different strategies worked (or didn’t) within those intervention settings. [269] While none of the studies are realist themselves, many pay close attention to context, process and interaction, so there is some rich information with which to start developing tentative hypotheses about how and why the interventions had the effects they did. [269, 290]

**Table 4.1. Definition and identification of concepts in SCMO configurations**

<table>
<thead>
<tr>
<th>Concept</th>
<th>Definition of concept in this review</th>
<th>How this was identified in the analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention strategy</td>
<td>Intervention strategies work by changing participants’ reasoning and resources. Importantly, they do not ‘work’ in a void, but interact with contextual features to generate change.</td>
<td>The intervention strategies listed in the results were identified from authors’ descriptions of attempts to support or advance policymakers’ capacity to use research in their work. These are listed by study in Appendix 4.</td>
</tr>
<tr>
<td>Context</td>
<td>Context is any condition that affects how people respond to intervention strategies (i.e. if and how mechanisms are activated). They include settings, structures, circumstances and the attributes and attitudes of those delivering and receiving the intervention.</td>
<td>Contextual features were identified primarily from authors’ accounts of intervention settings and circumstances before and during implementation. On occasion, they were inferred from information about responses to the intervention. In the SCMO tables that follow, we focus on aspects of context that relate specifically to each mechanism. A more general overview of context is provided first.</td>
</tr>
<tr>
<td>Mechanism</td>
<td>Mechanisms are how an intervention works. They are responses to the resources, opportunities or challenges offered by the intervention. Mechanisms are activated to varying extents (or not at all) depending on interactions between intervention strategies and contexts. Although mechanisms are not observable, their impacts are, so they can be inferred, tested and refined.</td>
<td>With one arguable exception [291], none of the studies explicitly identified or tested causal mechanisms so we inferred mechanisms from authors’ accounts of how the intervention was conceived, designed, delivered and received (i.e. how it was meant to function and how it actually functioned). This was supplemented with similar information from some of the non-eligible studies that were identified in this search, and from the wider theoretical and implementation literature. Mechanisms posited in the results tables that follow include hypotheses about: a.) How each intervention strategy worked (where mechanisms were activated successfully) and b) How strategies <em>would have worked</em> if they had been appropriate in that context (where mechanisms were not activated and their absence <em>may</em> account for poor outcomes).</td>
</tr>
<tr>
<td>Outcome</td>
<td>These are intended or unintended impacts generated by mechanisms. As described in the previous section, in this review outcomes may be proximal (process effects) or more distal study outcomes.</td>
<td>Outcomes of interest were explicit in most of the reviewed studies. Where they were vague, we inferred them from the studies’ research questions, interview foci and reported results. They are described in Appendix 4.</td>
</tr>
</tbody>
</table>

These terms are used in accordance with the realist movement associated with the UK’s RAMESES projects which aim to *produce quality and publication standards for realist research*. [292]

Following Best et al. [293] a six-step process was used in which we:

1. Read and reread the 22 articles to gain familiarity with the data.
2. Extracted details about what the intervention comprised, contextual features and the studies’ empirical findings, including any clues about interactions and causality.
3. Extracted program theory made by authors of these studies, i.e. explicit or inferred hypotheses, concepts or principles about what the intervention was expected to do or how it was expected to work. [269]

4. Reviewed the intervention strategies, findings and theories of related intervention studies that had similar aims but did not meet our inclusion criteria in order to harvest ideas about how those interventions had functioned.

5. Identified further relevant literature that might help develop new theories and/or additional detail. This included papers cited by authors of the primary studies and literature from fields that were likely to shed light on or challenge our findings. Time and resource constraints limited this step, so it was pragmatic rather comprehensive, drawing on materials known to the authors as well as found articles.

6. Summarised the connections we inferred between intervention strategies, implementation contexts, underlying casual mechanisms and observed outcomes in strategy+context+mechanism=outcome (SCMO) configurations. This was an iterative process that repeatedly cycled back through the activities described above. Table 4.1 describes how the main concepts were defined and identified in this process.

AH led this process. SB read the synthesis of studies and critiqued the draft SCMO configurations. SJR independently read the included studies, made notes in relation to steps 2 and 6, and critiqued the revised draft SCMO configurations. AH and SJR iteratively workshopped draft findings to reach agreement, drawing on feedback from our co-authors to resolve remaining areas of debate and to refine the final SCMO configurations.

Results

Search results

As Figure 4.2. shows, five, 14 and 18 articles respectively were identified from the three reviews. Of these 37 articles, 15 met our eligibility criteria. Another three articles were identified from the PAIS search, and three from the WoS search. A further article was identified from citation snowballing, resulting in 22 included studies. One article from the PAIS search was excluded on quality criteria: it provided so little detail of the intervention strategies and evaluation methods that we could not see how the conclusions were reached. Appendix 4 presents a tabular overview of the included studies’ aims, design, intervention strategies, participant numbers and characteristics, context, evaluation methods, outcome measures and key findings. This table includes theories, models or frameworks mentioned in the article as informing the design of the intervention or evaluation.

Study design

Study design terminology is often inconsistent. Here we use three terms to describe the overarching study design: Experimental = the research team provided the intervention and evaluated it using some form of randomisation and control groups, Interventional = the research team provided the intervention and evaluated it, but not using experimental methods, and Observational = the intervention or initiative being evaluated was not designed as part of a research study. Based on these definitions, 12 of the studies appeared to be observational, seven were interventional, and three experimental. These included process evaluations and opportunistic evaluations of projects, services or strategies that had been initiated by others.

Outcomes of interest were diverse, ranging from activities that can be measured objectively, such as increased use of systematic reviews [287], to more conceptual outcomes such as greater strategic use of data [291, 294], “catalysing” research use [129] and fostering a culture of critical thinking. [150]
Domains of capacity building and support

Eighteen strategies for supporting or increasing capacity were identified within four non-exclusive domains of research use capacity: Access to research (16 studies); Skills improvement in accessing, appraising and/or applying research in policy work (13 studies); Systems improvement tackling organisational or cross-organisational infrastructure, processes and/or resources (8 studies); and Interaction with researchers (11 studies). (Table 4.2.)

Most evaluations used case study methodologies. The main data collection methods were interviews (14 studies), focus groups (6 studies) and questionnaires (11 studies: 6 cross-sectional post-intervention and 5 pre/post). Outcomes, therefore, were largely self-reported. Two studies reported on the validity of their survey instruments. Eight studies also reviewed relevant documents, two conducted a network analysis, and one used observations of the intervention activities. Three studies used independent experts to assess intervention outputs. Participation data such as attendance rates at workshops and numbers of participants in the intervention and evaluation were reported to varying extents. The nine studies that included statistical analyses had diverse outcome measures, and used different sampling, data collection and modelling methods. See Appendix 4 for further details on aspects of study design.

Table 4.2. Research utilisation domains and strategies used within the reviewed studies

<table>
<thead>
<tr>
<th>Research utilisation domain</th>
<th>Intervention strategies (and no. of studies that used it)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access to research</td>
<td>1. Providing access to research articles or syntheses via an online database (5)</td>
</tr>
<tr>
<td></td>
<td>2. Disseminating tailored syntheses summaries or reports, including policy briefs (7)</td>
</tr>
<tr>
<td></td>
<td>3. Commissioning research and reviews (2)</td>
</tr>
<tr>
<td></td>
<td>4. Seminars or other forums in which research findings are presented (4)</td>
</tr>
<tr>
<td></td>
<td>5. Facilitated access using a knowledge broker (KB) or other intermediary (3)</td>
</tr>
<tr>
<td>Skills improvement</td>
<td>6. Skills development workshops (9)</td>
</tr>
<tr>
<td></td>
<td>7. Intensive skills training programs (5)</td>
</tr>
<tr>
<td></td>
<td>8. Training or support for managers in championing and modelling research use (4)</td>
</tr>
<tr>
<td></td>
<td>9. Mentoring (includes using knowledge brokers to build skills) (7)</td>
</tr>
<tr>
<td></td>
<td>10. Goal-orientated mentoring (with presentations or assessment) (4)</td>
</tr>
<tr>
<td>Systems improvement</td>
<td>11. Improving infrastructure e.g. library, new research portals, data sharing software (5)</td>
</tr>
<tr>
<td></td>
<td>12. Improving organisational tools, resources and processes e.g. procedures, toolkits, knowledge management protocols, funds for commissioning research (2)</td>
</tr>
<tr>
<td></td>
<td>13. Workforce development e.g. research-related positions and incentives (1)</td>
</tr>
<tr>
<td></td>
<td>14. Establishing internal research support bodies e.g. research units and committees (3)</td>
</tr>
<tr>
<td>Interaction</td>
<td>15. One-off or periodic interactive forums e.g. roundtables, cross-sector retreats, policy dialogues (4)</td>
</tr>
<tr>
<td></td>
<td>16. Platforms for ongoing interactivity e.g. community of practice, cross-sector committees (4)</td>
</tr>
<tr>
<td></td>
<td>17. Collaboration in the development of a research report or policy brief/dialogue (2)</td>
</tr>
<tr>
<td></td>
<td>18. Partnership projects: research co-production (3)</td>
</tr>
</tbody>
</table>

Two whole-of-intervention design strategies were also widely used: needs-based tailoring (10 studies) and multi-component programs (17 studies) (Table 4.8).
Table 4.3. Focus of intervention studies targeting research utilisation in policymaking 2001-2016

<table>
<thead>
<tr>
<th>Study reference</th>
<th>Intervention strategies as listed in Table 4.2</th>
<th>Intervention domains</th>
<th>Intervention design</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Access</td>
<td>Skills</td>
</tr>
<tr>
<td>1. Dobbins et al. 2001 [287]</td>
<td>2</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>2. Pappaioanou et al. 2003 [294]</td>
<td>1, 7, 8, 9, 11, 12</td>
<td>✓ ✓ ✓ ✓</td>
<td>✓</td>
</tr>
<tr>
<td>5. Wehrens et al. 2010 [296]</td>
<td>4, 18</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>8. Rolle et al. 2011 [291]</td>
<td>7, 8, 9, 10</td>
<td>✓ ✓ ✓</td>
<td>✓</td>
</tr>
<tr>
<td>9. Peirson et al. 2012 [150]</td>
<td>3, 6, 11, 12, 13, 14</td>
<td>✓ ✓ ✓ ✓</td>
<td>✓</td>
</tr>
<tr>
<td>11. Dagenais et al. 2013 [299]</td>
<td>2, 4, 15</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>12. Hoeijmakers et al. 2013 [300]</td>
<td>6, 11, 18</td>
<td>✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>13. Waqa et al. 2013 [301]</td>
<td>1, 6, 9, 10, 11</td>
<td>✓ ✓ ✓ ✓</td>
<td>✓</td>
</tr>
<tr>
<td>15. Traynor et al. 2014 [302]</td>
<td>1, 2, 5, 6, 8, 9, 12</td>
<td>✓ ✓ ✓</td>
<td>✓</td>
</tr>
<tr>
<td>16. Brennan et al. 2015 [212]</td>
<td>1, 6, 16</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>17. Dwan et al. 2015 [303]</td>
<td>4, 15</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>18. Shroff et al. 2015 [129]</td>
<td>2, 7, 14, 15</td>
<td>✓ ✓ ✓ ✓</td>
<td>✓</td>
</tr>
<tr>
<td>19. Uneke et al. 2015a [304]</td>
<td>6, 9, 10</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>20. Uneke et al. 2015b [305]</td>
<td>6, 7, 8, 9, 10, 16, 17</td>
<td>✓ ✓ ✓</td>
<td>✓</td>
</tr>
<tr>
<td>22. Langlois et al. 2016 [306]</td>
<td>7, 9, 16</td>
<td>✓ ✓ ✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
**Intervention participants and settings**

All studies targeted capacity in policymakers and/or policy agencies in that they were attempting to support, increase or otherwise improve: policymakers’ access to research; policymakers’ skills in accessing and/or using research; the capacity of systems in policy organisations to support research use; and/or interactions between researchers and policymakers that were intended to facilitate knowledge exchange or partnership work. Many studies had more than one category of participant, e.g. a mix of policymakers, practitioners and researchers. The majority included bureaucrats in government departments of health or equivalent at the regional level (11 studies) or national/international level (9 studies). Eleven studies targeted government employees running regional health services, and two included elected (ministerial) policymakers.

The intervention settings spanned 18 countries. There were 17 single-country studies conducted in: Canada (5), Australia (3), Nigeria (3), The Netherlands (2), Burkina Faso (1), Ethiopia (1), Fiji (1) and the USA (1). Four were multi-country studies conducted respectively in Bangladesh, Gambia, India and Nigeria; Cameroon and South Africa; Bolivia, Cameroon, Mexico and the Philippines; and Argentina, Bangladesh, Cameroon, Nigeria and Zambia. And one was an international collaboration run from Canada. Thus 10 studies took place within one or more low- and middle-income countries. [307]

**Program theories**

The 22 studies draw on a diverse suite of theories and concepts. None present a formal program theory, but many use frameworks to guide intervention development. The RCT conducted by Dobbins et al. [295] is based on diffusion of innovations [11, 12, 308]; Brennan et al. [212] use the theoretical domains framework [309]; and Shroff et al. [129] adapt a knowledge translation framework. [249] Others draw eclectically on concepts from a variety of sources including:

- studies of how research is used in policymaking—both systematic reviews [145, 310] and individual studies such as Weiss’s seminal typology of research utilisation [85]
- models for mobilising research such as knowledge transfer frameworks [311-313] and partnership approaches [214, 263, 314-316]
- analyses of barriers to researcher-policymaker relationships [78, 80] and ‘gap-bridging’ solutions such as the linkage and exchange model [317], and the use of knowledge brokers [318]
- studies of organisational support for research use [259, 319]
- guidance for facilitating the use of research in policymaking, including in low- and middle-income countries, e.g. the SUPPORT tools developed by Oxman, Lavis et al. [320-322]
- WHO commissioned reports on building capacity for health policy and systems research. [253, 256]

A minority of studies developed their own program theory or conceptual framework that guided the intervention design and evaluation [e.g. 283, 294, 300], and some report using frameworks primarily for the evaluation. [e.g. 150, 247, 291]

**Intervention strategies, contexts, causal mechanisms and outcomes**

The findings derived from our realist analysis of the 22 studies are now presented. See Appendix 4 for a summary of each study’s design, outcomes and informing theory.

**Overarching contextual considerations and their implications**

Nearly all the studies in this review conceptualised research use as contextually contingent. They assumed that some degree of responsivity to local needs was required by research providers, and that policymakers’ judgements about the usefulness of research were flexible, according to shifting circumstances, and based on far broader criteria than academic hierarchies of evidence, e.g. “Research is only as useful as potential
users perceive it to be, irrespective of its methodological rigour or its findings' power". [303:241] Some pointed to the limitations of technical-rational models of research use in political decision-making [287, 294, 303], and over half emphasised the complexity of policymaking. [129, 150, 212, 236, 247, 261, 294, 295, 299, 300, 304, 306] Terminology reflected the acceptance that policy will never be based entirely on research. [61] Indeed, few of the studies used the term evidence based policy unquestioningly, preferring more nuanced post-evidence based terms such as “evidence-informed policy” [129, 298], “evidence-informed decision making” [150, 295, 302], and “research-informed policy” [212, 236, 301].

From our analysis, it appeared that there were some similar contextual factors in all the studies reviewed, despite their very different settings (e.g. Burkina Faso and Canada). This suggests there may be universal influences on the use of research in policymaking of which virtually every capacity building intervention should take account. The main contextual factors identified were:

**Research characteristics:** Policymakers’ use of research was influenced by the degree to which they were able to obtain research that was relevant, applicable and easy to read. This suggests the need for strategies that increase the availability and accessibility of fit for purpose research findings. Credibility of research and researchers is also a consideration.

**Individual characteristics:** Policymakers’ use of research was affected by existing research-related knowledge, skills and self-confidence; and views about the value of research. The former clearly indicates task-appropriate skills development and support, but the latter suggests that, in some cases, it will be necessary to influence beliefs.

**Interpersonal characteristics:** Although neither community is heterogeneous, there were common differences between policymakers and researchers in terms of language, values, expectations and incentives. This suggests the need for strategies that either bridge these communities, or form more cohesive connections between them, potentially blurring their boundaries.

**Organisational characteristics:** Research use in policy agencies was shaped by organisational culture. Agency remits, resources and constraints further influenced how research was prioritised and used. This suggests the need to activate structural mechanisms that increase expectations of, and facilitate, research use in day-to-day practice. Underlying values and assumptions may need to be influenced. Leadership by managers and opinion leaders is likely to be key.

**Environmental characteristics:** Policymaking environments were complex, political and responsive, affecting what research could be used for what purposes, and the time available for this. The way that research is (or can be) interpreted in this argumentative arena is likely to determine its role in policy processes, thus relevance, applicability and credibility are not fixed research characteristics but determined in relation to circumstances. This suggests that tailored research (both in terms of methodology and presentation), rapid production of findings, and responsive dialogue with researchers may be valuable. Methods for supporting this are likely to include commissioning and/or research-policy partnerships and/or internal generation of research by policy agencies.

These overarching contextual factors align with other reviews, which conclude that policymakers’ capacity and motivation to use research is shaped by forces such as these at micro, meso and macro levels. [145, 248, 310]

The next four sections present our findings in relation to the four domains of capacity previously identified. As per the focus of this review, the emphasis in these results is not on the extent to which interventions were successful in effecting change, but on how change was effected or why it was not; consequently, the narrative overview in each of the results sections is on mechanisms. The tables that follow place these mechanisms in context by showing: what intervention strategy was used; key contextual factors identified in
the studies; possible causal mechanisms including those that appear to have been activated in the studies and those that were apparently required but were not activated; and any reported results relating to that strategy (including process effects and study outcomes). [15, 323] Our hypotheses describe mechanisms that were inferred from multiple studies and/or powerfully evident in one study, and are supported by theoretical or empirical implementation literature. Where low effects are observed, we hypothesise that one or more key mechanisms were not activated, or were activated too weakly to bring about the level of desired change. Note that the contextual factors described above are considered to be ‘givens’ in these tables and so are only reiterated where they seem to be most crucial.

**Access to research**

Mechanisms that appeared to underpin access to online research included awareness of resources and the relative merits of different kinds of research within them, valuing what was on offer, the efficiency with which research could be obtained, and confidence in using resources and their contents. When research was synthesised, tailored for specific users, and sent to them, ease of access and ease of use aided uptake, probably aided by increased policy-relevance and applicability. As with all domains, perceived fit between what was offered and policy needs/priorities was key. The tailored, contextualised and inclusive evidence synthesis provided by evidence briefs tick many of these boxes. Commissioning rapid reviews maximised policymakers’ engagement with and control over the research focus, methods and timeliness. The costs and process of commissioning are likely to increase investment in using the end-product.

The value of seminars seemed to be enhanced by tailoring and interactivity, and by the credibility and communicative skills of the presenters who engage policymakers despite often dry content. Meeting researchers at these seminars can break the ice and lead to further interaction.

Intermediaries such as knowledge brokers can facilitate access to research by providing navigational support in unfamiliar terrain. They provide a communicative bridge by helping policymakers articulate needs and expectations and, in some cases, translate these for researchers. The intermediaries’ interpersonal skills, credibility, availability, ability to provide individualised support and perceived neutrality enabled the relationship to work, but this also requires time in less research-orientated settings.
Table 4.4. *Access to research* intervention strategies, context, mechanisms and impacts

<table>
<thead>
<tr>
<th>Intervention strategies used</th>
<th>Contextual factors that affected how these strategies worked</th>
<th>Hypothesised mechanisms (how did, or why didn’t, the strategies work in these contexts?)</th>
<th>Observed process effects &amp; outcomes of studies using this strategy*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Providing access to research articles or syntheses via an online database [212, 295, 301, 302]</td>
<td>Research characteristics: • Availability of policy-relevant, applicable and accessible research</td>
<td>• Awareness of the resource and of what the research content offers  • Valuing what is offered  • Efficiency – portals and syntheses maximise access to some research  • Perceived fit between what is offered and policy needs/priorities  • Confidence in accessing and using resources and contents</td>
<td>• One study found poor awareness of availability and limited post intervention appreciation of the ‘value’ of systematic reviews  • Two studies found low use of registries of systematic reviews when provided alone or in conjunction with other intervention strategies  • Overall, the studies indicated that improved access alone does not significantly improve research-informed decision making</td>
</tr>
<tr>
<td>Disseminating tailored syntheses summaries or reports, including policy briefs [129, 212, 283, 287, 295, 297, 299, 302]</td>
<td>Individual factors: • Awareness of where and how to access research  • Awareness of the scope or potential usefulness of what is available  • Existing level of experience/expertise and confidence in using research</td>
<td>• Needs-based tailoring increases policy-relevance and applicability  • Ease of access – research comes to you  • Ease of use – concise, clearly presented ‘scannable’ findings which have been quality-checked and compiled by experts so are trustworthy and overcome the barrier of limited critical appraisal skills  • Evidence contextualisation and integration (policy briefs)</td>
<td>• Tailored research products were well received in five studies, e.g. perceived as credible, useful and likely to impact decisions  • Varied outcomes re use and influence, but largely positive e.g. results of single studies showed:  - Tailored systematic reviews used to influence decision making by 63% of policy makers  - Tailored reports used to confirm policies, e.g. to check current practices against the evidence  - Tailored information (in addition to access) increased research-informed policies significantly more than increased access alone</td>
</tr>
<tr>
<td>Commissioning research and reviews [150, 236]</td>
<td>Organisational factors: • Extent to which research use is valued and expected in the organisation  • Extent to which means of accessing and supporting research use (people, resources, systems) are in place  • Level of leadership around research use  • Degree to which strategies and resources for increased access can be embedded within the organisation</td>
<td>• Engagement with and control of the research aids relevance and applicability  • Investment in using the research  • Timeliness – rapid review processes decrease lag time</td>
<td>• In one study, commissioned reviews facilitated by a knowledge broker were:  - Seen as useful by policymakers who commissioned them  - Considered to be accurate  - Used in mostly indirect ways, e.g. informing policy deliberations and providing background information</td>
</tr>
<tr>
<td>Seminars or other forums in which research findings are presented [247, 296, 299, 303]</td>
<td>Environmental factors: • Current policy/political circumstances affecting the amenability for research use, including governmental commitment to research-informed policy</td>
<td>• Policy-relevant, applicable and accessible content – aided by tailoring and interactivity  • The presenters’ perceived credibility and skills  • Breaking the ice – interaction can lead to relationship development</td>
<td>• Well received by attendees in four studies, e.g. relevant and better than reading reports  • Little discernible behaviour change in two studies  • Poor onwards dissemination within policy organisations in one study</td>
</tr>
</tbody>
</table>
| Facilitated access using a knowledge broker (KB) or other intermediary [236, 295, 302] | **Interpersonal factors:**  
- Availability of researchers and other experts  
- Levels of actual and anticipated understanding and respect between researchers and policymakers (these interpersonal factors relate to the last three strategies only)  
- Understanding of KB role and support available (crucial for engaging with KB)  
- Bridging – help articulating needs and expectations across research-policy boundaries  
- Navigational support in developing strategies and using systems  
- Relationship development founded on:  
  - trust and rapport, which needs KBs’ interpersonal skills & credibility  
  - time to form relationship and start seeing positive results  
  - independence/neutrality (in one study)  
- Individualised support and assurance - KBs can guide and confirm research practices  
- Availability of KB ‘on demand’ maintains momentum in work and learning, reducing stress | **KBs regarded as helpful and preferable to training and tools (two studies)**  
- One study found no benefit when KB added to other strategies (e.g. the use of KB + tailored summaries)  
- Working closely with KB resulted in increased knowledge and skills in finding, appraising and using evidence, and confidence in applying these skills and engaging in research informed decision making (two studies)  
- When working with KB to develop knowledge products, those products were perceived as useful and accurate, and were used (two studies) |  
* In this and subsequent tables, not all the studies that target each domain will necessarily be included, for example, where a study’s strategies for increasing access was based on skills- or systems-improvement, or interaction, it is not cited in the table above. Calculations of the number of studies where process effects/outcomes were observed (in the last column) are based only on studies cited in the intervention strategies column. |

**Skills improvement**

Mechanisms for skills improvement in using research appear to include: policymakers believing in the relative advantage of participation which is affected by the perceived appropriateness/desirability of intervention goals and the relevance, applicability, accessibility and credibility of intervention content. *Andragogical principles* that emphasise participant-driven learning (manifested in a partnership approach that may include needs consultation, tailored content, informal information exchange and practice opportunities) engages policymakers. Participants’ active input appears to maintain interest and investment. Strengths-based learning that develops self-efficacy increases motivation and can empower policymakers to become research champions and train or mentor others. Strong leadership support for the intervention and its goals, including *modelling* research use practices, is emblematic of wider organisation commitment to using research. Targeted policymakers will have to find training manageable if they are to attend; this may require pragmatic timing and workarounds.

Mentoring works by providing individualised guidance and support about the real-world application of new knowledge and skills, which in turn increases self-efficacy as abstract learning is turned into concrete practice. Mentors’ credibility and experience, and relationship skills, are crucial. Participants’ accountability, triggered by the need to present their work and/or have it assessed, increases motivation to develop competence in using new knowledge and skills.
Table 4.5. Skills improvement intervention strategies, context, mechanisms and impacts

<table>
<thead>
<tr>
<th>Intervention strategies used</th>
<th>Contextual factors that affected how these strategies worked</th>
<th>Hypothesised mechanisms (how did, or why didn’t, the strategies work in these contexts?)</th>
<th>Observed process effects &amp; outcomes of studies using this strategy*</th>
</tr>
</thead>
</table>
| Skills development workshops [150, 212, 247, 298, 300-302, 304, 305] | Individual factors:  
  • Existing skills and confidence in accessing, appraising and using research  
  • Perceptions of policymaking as an applied practice |  
  • Relative advantage – perceived value of new skills over current skills  
  • Andragogy – participant-driven learning inspires interest and investment in learning  
  • Self-efficacy – strengths-based learning motivates and assures policymakers of their capabilities  
  • Training empowers policymakers to use and champion research  
  • Leadership and modelling – managers demonstrate and advance organisational commitment to a research-oriented culture  
  • Manageability of training with busy workloads (flexible dates may help)  
  • Workshops create an enabling environment for interaction between policy makers and experts |  
  • Attendance variable: ranged between 65% to 87% in studies that reported on it  
  • Workshops and training well received by attendees  
  • Shorter, more intensive programs associated with increased retention in training  
  • Sustainability undermined by high staff turnover  
  • Self-reported increases in knowledge, skills & confidence to:  
    - access, interpret, synthesise and apply research  
    - work in cross-sector partnerships  
    - identify and respond to critical health problems  
    - promote research and research-informed policy  
    - develop research-informed policy materials  
  • Improved relationships with researchers (less distrust)  
  • Training alone had intangible effects on policymaking |
| Intensive skills training programs [129, 291, 294, 305, 306] | Interpersonal factors:  
  • Existing relationships (and communication) between policymakers and researchers, including disciplinary silos and level of trust  
  • Degree of enthusiasm about closer working ties (only relevant to strategies that require close working relationships)  
  • Organisational factors:  
    • Existing organisational remit and support for research  
    • Staffing characteristics, e.g. personnel shortages, levels of staff turnover  
  • The extent to which managers and general staff agree on current capacity and need for further development |  
  • Project- and person-specific support helps to develop practical skills and increase self-efficacy  
  • Credibility of mentors engendered by applied expertise and interpersonal skills  
  • Concretisation – applying new skills turns theory into practice  
  • Accountability – participants strive to integrate new skills to demonstrate competence |  
  • Participants applied skills in practice  
  • Self-reported increase of applied skills and confidence  
  • Some participants went on to train others  
  • Contributed to a culture of continuous learning and sharing  
  ... in addition to the above...  
  • Expert assessment reported increased standards of practice, but room for further improvement  
  • Higher rates of completion when mentors had policy rather than academic expertise |
| Training or support for managers in championing and modelling research use [291, 294, 302, 305] | Environmental factors:  
  • Competing external pressures and opportunities, e.g. disease outbreak |  
  •  |
| Mentoring (includes using knowledge brokers to build skills) [291, 294, 301, 302, 304-306] |  |
| Goal-orientated mentoring (with presentations or assessment) [291, 301, 304, 305] |  |

Systems improvement

The mechanisms underpinning systems improvement appear to be diverse, reflecting the breadth of strategies. Diffusion of innovations theory [11, 12, 324] helps make sense of findings across the studies in relation to interactions between new infrastructure, tools and processes. It posits that new systems must be:
compatible with key professional and organisational culture and values, flexible enough to accommodate aspects of existing practice that participants will not relinquish, sufficiently easy to use so that policymakers are not deterred, and have relative advantage, i.e. seem better than existing systems for the individual and, probably, better than existing systems for the organisation, so it feels worth the effort of adaptation. Participatory planning and implementation of systems improvement with potential participants may most effectively engage and enthuse them, increasing their readiness for change as well as making the intervention more fit for purpose.

Improved systems widen opportunities for using research by increasing ease of access. Where research skills are brought into and developed within the organisation, there is strengthened belief in managers’ commitment to research use. Co-location and control of expertise is likely to increase the policy-relevance, applicability, accessibility and, probably, timeliness of research outputs and advice. In-house research expertise provides opportunities and incentives that policymakers may find motivating. In general, systems improvements help to embed research use in day-to-day practice and demonstrate managerial commitment, both of which contribute to a research-oriented culture.

Table 4.6. Systems improvement intervention strategies, context, mechanisms and impacts

<table>
<thead>
<tr>
<th>Intervention strategies used</th>
<th>Contextual factors that affected how these strategies worked</th>
<th>Hypothesised mechanisms (how did, or why didn’t, the strategies work in these contexts?)</th>
<th>Observed process effects &amp; outcomes of studies using this strategy*</th>
</tr>
</thead>
</table>
| Improving infrastructure, e.g. library, new research portals, data sharing software [150, 247, 294, 300, 301] | Research characteristics:  
• Availability of policy-relevant, applicable and accessible research  
Individual factors:  
• Beliefs about the value of research  
• Skills and confidence in using research-related systems  
Interpersonal factors:  
• Quality of relationships between research support staff and policymakers  
Organisational factors:  
• Organisational culture and practice norms, e.g. the value placed on tacit versus scientific knowledge  
• Existing systems for supporting research use  
• If and how systems advantage or disadvantage members of staff  
Environmental factors:  
• Institutional norms and change trajectories in the sector re research use and systems | • Compatibility of new system with organisational values  
• Flexibility of new system to accommodate aspects of existing practice  
• Perceived relative advantage of new system including its ease of use, make change worth the effort  
• Participatory development and implementation of system increases utility, and engages and enthuses recipients  
• Strengthened belief in managerial commitment to research – organisational systems have a functional and a representational role: they serve the organisation’s routine practices but also signal its values [258] | When used in conjunction with training and other supports (e.g. knowledge brokers, mentors), systems improvements were believed to contribute to:  
• improved capacity to access, understand and use research  
• improved capacity to identify and address health problems using research  
• increases in the amount of research syntheses produced internally  
• improvement in the rigour of research syntheses produced internally  
• successful partnership project work  
• growth of a research-oriented organisational culture |
| Improving organisational tools, resources and processes, e.g. procedures, toolkits, knowledge management protocols, funds for commissioning research [150, 294, 302] | Workforce development, e.g. research-related positions and incentives [150] |
| Establishing internal research support bodies, e.g. research units and committees [129, 150, 247] | Establishing in-house and committees e.g. research units and support bodies |
Interaction with researchers

Mechanisms for productive interactions between policymakers and researchers appear to include mutual commitment to investing time and effort in interaction, and mutual interest (including the identification of benefits to both parties) in the endeavour. Trust, respect and communicative ease underpin relationship formation, but this takes time to develop and, may require repeated interactions. Also, that the researchers are perceived as neutral, dispassionate contributors.

Where positive interaction is underway it can sensitise and upskill both parties via learning from the ‘other’ about their values, work contexts and practices. Interactions are more sustainable when there is strong organisational support, and where formal arrangements are put in place rather than relying on individuals (who may move on). Leadership and championing from respected ‘insiders’ may motivate staff to engage with the intervention and put it into practice. Known contacts can act as linkage agents, introducing people to networks and keeping them connected. Collaboration increases ownership of and investment in the research process and outputs, but only when it is genuine, i.e. when both parties have the power and ability to shape critical decisions and have input into processes. However, genuine collaboration is often hard to facilitate. Good governance arrangements can help by ensuring that costs and rewards are agreed and shared, roles are clear and expectations are articulated and met. Reflexivity—paying attention to partnership processes, critiquing and seeking to learn from them, perhaps through developmental evaluation approaches—may combat the lure of traditional silos and disciplinary norms that have been found to undermine collaborations.
### Table 4.7. Interaction intervention strategies, context, mechanisms and impacts

<table>
<thead>
<tr>
<th>Intervention strategies used</th>
<th>Contextual factors that affected how these strategies worked</th>
<th>Hypothesised mechanisms (how did, or why didn’t, the strategies work in these contexts?)</th>
<th>Observed process effects &amp; outcomes of studies using this strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-off or periodic interactive forums e.g. roundtables, cross-sector retreats, policy dialogues [129, 247, 294, 299, 303]</td>
<td>• Degree to which the parties have different norms and expectations (may be linked to disciplinary differences) • Perceptions of the value of interaction / collaboration</td>
<td>• Mutual commitment to interaction • Mutual interest in the topic or research issue • Trust and perceived mutual respect • Perceived neutrality of research partners (not pushing their own agenda) • Learning from the ‘other’ about values, work contexts and practices</td>
<td>• Uneven attendance, especially at senior levels • Well received: relevant and stimulating and perceived to have achieved goals • Self-reported “broadened knowledge” and improved relationships with researchers • No tangible practice impacts</td>
</tr>
<tr>
<td>Platforms for ongoing interactivity e.g. community of practice, formal networks, cross-sector committees [212, 261, 305, 306]</td>
<td>• ... in addition to the above... • Repeated face-to-face contact allows for development of: trust, respect and communicative ease • Linkage – known contacts link partners to networks and keep them connected • Engaged leadership/championing by ‘insider’ opinion leaders re the initiative and/or research use • Organisational support can help to initiate and sustain interactions • Genuine collaboration at all stages increases ownership of and investment in the research process and outputs • Goal alignment across all parties • Formalisation of arrangements, especially good governance can ensure: - mutual benefits are identified - costs and rewards are shared - roles are clear - expectations are articulated and met - concerns are raised and conflicts are managed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaboration in the development of a research report, evidence brief or policy dialogue [283, 305]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partnership projects: research co-production [261, 296, 300]</td>
<td>• Time available to engage, interact and collaborate</td>
<td>• Adaptivity helps approaches stay fit for purpose within a shifting policy context • Reflexivity – attending to and learning from networked and partnership processes</td>
<td>• Networked activities were valued, but two studies reported poor awareness of the initiative’s purpose and resources or projects • Three reported the establishment of more trusting and equal relationships • One reported the lack of a ‘common language’ • Some members acted as ‘information conduits’ linking the network to their own organisations • There was some self-reported increase in understanding of research use in policy • Only one study reported any policy impacts</td>
</tr>
</tbody>
</table>
One of the challenges in evaluating interactive initiatives is the increased entanglement of strategies, mechanisms, process effects and outcomes. For instance, existing positive cross-sector relationships may function as both a context and a mechanism; trust may function as both a mechanism and process effect; and improved relationships may be an outcome while also providing context for further dialogue and partnership work. Thus, there are dual functions and feedback loops implied in much of this theorising.

**Whole-of-intervention design strategies**

Although many interventions were described as “tailored”, only 10 of the reviewed studies both reported using formal needs analyses or consultative/collaborative strategies for determining needs and preferences and gave some indication of how this shaped the intervention. There was very little information about how this might have impacted responses to the intervention. Nevertheless, it seems likely that tailoring based on accurate needs assessment will maximise the interventions’ compatibility with local needs and practices, and its ability to build on local strengths. Where participants collaborate in tailoring they are more likely feel like respected partners in the intervention and thus to have ownership of and investment in its outcomes.

As shown in Table 4.3, most studies that employed multiple intervention strategies did so to improve capacity across two or more domains (e.g. access, skills and interaction) in order to address different levels of support and constraint in research use. Only three studies used multiple intervention strategies to improve capacity in a single domain (e.g. a combination of training workshops, mentoring and practice assessment were used in conjunction to build individual skills). Possible mechanisms are triggered by the interaction and complementarity of multiple strategies, which may be increased when multiple domains are targeted because strengthening capacity in one area (e.g. organisational systems) is likely to support capacity growth in other areas (e.g. individual skills); and strategies may function synergistically to shape a conducive environment for research use. As such, they may both represent and facilitate a culture of research use.

**Table 4.8. Key whole-of-intervention strategies, context, mechanisms and impacts**

<table>
<thead>
<tr>
<th>Intervention strategies</th>
<th>Contextual factors</th>
<th>Hypothesised mechanisms</th>
<th>Potential process effects/outcomes *</th>
</tr>
</thead>
</table>
| Local tailoring based on needs/situational analysis [247, 294, 301, 302, 306] | Each intervention site has unique features that interact with the implemented strategies [23]. Policymakers have existing strengths and skills. Their needs will vary. | • Tailoring based on high quality needs/situational analysis maximises intervention compatibility, including its ability to build on local strengths and tackle areas of real need  
• Where participants collaborate in tailoring they feel respected/heard and have increased ownership of and investment in the intervention outcomes | • Greater acceptance of the intervention’s local fit and utility  
• Active support of the intervention and its goals |
| Using multiple strategies to target different forms and levels of capacity [129, 150, 212, 247, 291, 294-296, 299-306] | Research use is multi-factorial. Capacity exists in different forms and at different levels (individual, interpersonal, organisational and wider environmental). Supports and constraints in one area or at one level affect responses in others | • Strengthened capacity in one area supports capacity growth in other areas  
• Strategies interact synergistically to shape a conducive environment for research use | • Greater, and more sustainable, change in using research |

* These outcomes are speculative: there was no clear evidence of outcomes relating to these strategies in the studies
Discussion

The 22 studies in this review display a diverse suite of theories, concepts and frameworks from different disciplines and fields. We identified 18 intervention strategies and two dominant design strategies targeting four domains of research use capacity: Access, Skills improvement, Systems improvement and Interaction. These studies reflect dominant concerns in the literature about the paucity of policy-usable research and difficulties locating it within information-saturated environments; the need for policymakers to have adequate skills and confidence in using research, and for work processes and infrastructure to support this use; and the benefits of researchers and policymakers developing relationships that facilitate mutual understanding, information exchange and collaboration. Underpinning much of the above, are concerns about how the value of research in policy processes is perceived by policymakers and how these beliefs are affected by organisational cultures.

Despite drawing on ideas from different traditions, most of these studies rejected linear pipeline models in which universally applicable research findings can be ‘transferred’, and favoured more nuanced notions of both the research product and the policy process. Prominent ideas included: policymaking as information bricolage in which research findings are only one component [e.g. in 129, 247, 294, 295, 305]; the rhetorical and political dimensions of research use [e.g. in 283, 296, 299, 303]; and research use as situated – dependent on myriad fluctuating contextual factors. [e.g. in 129, 212, 303] Correspondingly, the findings indicated scant instrumental use of research. Indeed, they showed that even where specific research was valued and understood it was seldom translated directly into policy action [236, 283], and that some forms of use did not correspond with established typologies. [283]

Like others, we cannot identify one strategy as superior to others in building the capacity of policymakers to use research. [247, 273] Policymaking is a complex and contingent process, and the various capabilities that facilitate it operate at multiple levels, including the meso and macro levels where local infrastructures, politics and issue polarisation is likely to impact what is viable. [247, 248] A combination of strategies that are responsive to changing conditions are likely to be most appropriate. And regardless of the design features of the intervention, it will be interpreted and enacted differently in different settings. [72] Nevertheless, there are lessons from the 22 studies in this review that have transferable implications for the design and implementation of research utilisation capacity building interventions in policy agencies, some of which we now discuss.

Access

It is axiomatic that policymakers cannot use research if they do not know about it. To this end, efficient routes to relevant, clearly presented research findings are a boon. Tailored and contextualised syntheses—either in document form or via presentations, seminars and advice from knowledge brokers or researchers—seem to offer the most helpful means of providing this access. The benefits of tailoring information for specific audiences and using active dissemination are supported by other reviews. [134]

While these studies clearly demonstrated the importance of research being policy-relevant, applicable and credible, these concepts raise problems of their own. Regarding credibility, participants did not always judge the merits of research using academic hierarchies of evidence. Weiss suggests this is because policymakers assess research credibility using ‘truth tests’ (are the findings plausible / legitimate?) and ‘utility tests’ (are proposed solutions feasible in our context?). [86] Consequently, local data is often most compelling, and contextualisation is needed for the findings to have leverage in the discursive and rhetorical processes that characterise policymaking. [61, 325] This suggests it may be unhelpful for interventions to focus solely on access to untailored systematic reviews and syntheses. Enhancing research for policy purposes involves trade-offs: increasing one attribute — relevance, credibility or accessibility — is likely to be at the expense of another. For example, presenting research findings clearly can enhance accessibility and relevance, but may
also neglect important complexities which decreases credibility. Therefore solutions may be most effective when tailored on a case-by-case basis. [326, 327]

Most of the interventions that attempted to increase access appeared to conceptualise access as necessary but insufficient for effective use of research, hence their parallel attempts to address individual, interpersonal and organisational capabilities. They recognise that research use is an intensely social and relational process [134], and that to increase it we have to understand and work with supporting factors such as organisational culture, professional behaviours, local circumstances and different intervention facilitation approaches. [13, 328-331]

**Skills**

Training workshops—the primary intervention for individual capacity building—appear to provide a useful starting point providing they are well-tailored (resulting in relevant and appropriately pitched content) and facilitate active input from participants. Workshops are generally well received with high levels of self-reported improvement in understanding, but as a stand-alone intervention method they seem unlikely to result in substantial practice change. Uneke et al. praise the merits of one-off workshop [298], but follow-ups of RCTs find workshops alone to be costly and largely ineffective [277]: without support structures and systems even the best training will not be translated and sustained in practice. (150, 233, 294) The trade-off between intervention intensity and attendance by busy policymakers—especially those at higher levels of seniority who also have a role in modeling and championing change—remains problematic.

The use of mentored practice seems to address some of these concerns, and is supported in the wider literature. The hypothesis that mentoring develops knowledge, skills and confidence has been tested in multiple studies with success where the mentor is appropriate and the mentor/mentee relationship is sound. [332, 333] Wider benefits include connecting mentees’ to communities of practice, and inspiring them to become mentors themselves. [334] Shroff et al. [129] suggest that, in policy agencies, this requires that the mentor has local knowledge and applied policy expertise. Others note difficulties in identifying mentors and matching them with mentees, and in freeing up mentors’ time sufficiently. [256, 335]

Combining training and mentoring with performance goals and assessment (as three of the reviewed studies did) may offer the best option for embedding skills. For example, in their multi-country study Pappaioanou et al. conclude that “without supportive follow-up and supervised application of skills, participants frequently continued to use the same work practices that they had used before they attended the training”. [294:1935] A recent meta-analysis found that goal-focused mentoring (otherwise known as coaching), even when short-term, improved individual, skills-based and affective outcomes. [336] Mentoring may offer greater support to staff who are less engaged in the workforce, so policymakers who are new employees and/or especially lack confidence in using research skills may benefit most. [335] The terminology in this area is muddled so it is important to consider the specific tactics and goals of the intervention rather than relying on terms such as knowledge brokering, coaching or mentoring to define them.

**Systems**

Knowledge utilisation is intimately linked to organisational structure and systems [31], so it is not surprising that these appear to play a key role in supporting individual efforts to access and use research. However, they must be fit for purpose, attuned to real practice needs, able to accommodate local adaptations and provide a clear benefit. Those developing and implementing such systems cannot afford to neglect the complex human dynamics within which they must work, consequently, participatory development of systems interventions may offer the best chance of success.
The outcomes of research-focused recruitment and performance management were not generally available in these studies, partly because their effects are often hard to disentangle from other strategies; however, they promise proximal and distal benefits. In-house research experts such as knowledge brokers may be more able to provide highly relevant, applicable, accessible and timely findings, but can also help to build wider capacity by supporting their colleagues’ skills development, and contribute to a more research-orientated organisational culture. Evaluations of knowledge brokering in Scottish government departments and in Canadian healthcare organisations show a positive impact on research use [337, 338], and their use has been found to strengthen the clarity of research commissioned by policymakers. [339] Our review found that the use of onsite knowledge brokers had mixed results, possibly because of the time needed to build productive working relationships with policy staff. [302] Thus longer-term use may be most beneficial. Findings concur with descriptive and other empirical studies about the importance of knowledge brokers’ interpersonal skills and credibility. [340-342]

Interaction

There is little doubt that interaction between policymakers and researchers—when it is positive and productive—tends to operate as a virtuous circle that increases trust and confidence in the benefits of further dialogue, and builds the capacity of both parties to understand and work with the other. For example, strategies that modelled respect for policymakers as ‘knowers’ as well as ‘doers’ may have increased engagement, e.g. having senior policymakers co-facilitate deliberative forums. [298] Mutual respect and commitment seemed to be crucial mechanisms, suggesting that those selected to take part in these initiatives should be carefully selected where possible, and that enthusiastic but sensitive facilitation might be helpful in the early stages. Results also suggest that reflexivity and continual adjustment may be crucial in dealing with the inevitable challenges of collaboration. There are tools available to help parties prepare for partnership work [e.g. 343], and to monitor and evaluate its functioning. [214, 261]

The extent to which interaction translates into research-informed policymaking is less certain. Neither increased understanding or collaborative outputs necessarily influence policy decision-making; however, where sound relationships are formed they do appear to support the ‘social life’ of research—helping findings and ideas from research move into, within and between networks. [92, 95, 344, 345] Empirical studies repeatedly find that professionals, including policymakers, are more likely to seek and use research obtained via trusted interpersonal channels rather than from formal sources. [31, 118, 346] Interaction can build relationships that enable researchers to operate within this sphere of trust and familiarity. [118, 260]

Despite disappointing outcomes in the three of the five collaboration-focused studies, co-production remains a worthy goal, particularly in the light of a recent review that found an association between clinician involvement in research and improved healthcare performance. [30] The sticking point appears to be the capacity of individuals and organisations to facilitate genuine collaboration in which roles and tasks, resources and outputs are negotiated, and leadership is distributed across boundaries, resulting in shared expectations and mutually satisfying returns on investment. Early robust dialogue and fair but firm governance arrangements, underpinned by institutional support, seem to play an important role. The extent to which these policymakers experienced a sense of ownership in the research process is likely to have been just as vital. As Zimmerman argues [1], ownership and buy-in are opposite concepts: ownership means collaborative development of ideas, decision-making and action; whereas buy-in means agreeing to someone else’s proposal. For example, in Kothari et al.’s intervention the policymakers’ involvement seems to have been limited to articulating the research questions and commenting on draft versions of the report. This consultative role places them closer to ‘endorsers’ than ‘co-researchers’ in the spectrum of co-production. [264] As Senge [347] puts it, people feel ownership of a shared vision not when they are playing according to the rules of the game, but when they feel responsible for the game. These findings align with other studies: a systematic review concluded that knowledge exchange in policymaking depends on the establishment of a viable cost-sharing equilibrium and institutionalised communication networks. [248]
Evaluations of the CLARHC partnerships concur and draw attention to the benefits of leveraging existing relationships when starting partnerships. [193, 237, 348]

Key considerations in intervention design

The lack of detail about needs/situation analysis and how it was used to tailor interventions makes it hard to draw conclusions about the mechanisms that were or weren’t triggered. However, others argue strongly that generalised capacity building interventions are seldom successful; rather, they should be designed in response to accurate analysis of existing capacity and concerns, research needs and local conditions, derived from consultation or—better still—collaboration with potential participants. [349] This is supported by calls more generally for collaborative needs assessment as a precursor to local tailoring of interventions [350] [351] as this is likely to identify goals that are locally meaningful, make implementation plans more actionable, actively engage participants in translating data, and increase their investment in outcomes. [352, 353] Understanding existing capacity is also vital for tapping into local practice strengths. [267] As Trostle et al. argue, “capacity can often be increased more effectively by reinforcing existing structures than by building new ones.” [354:63]

Findings emphasised the power of organisational culture to shape receptivity to intervention ideas and resources. This suggests that tailoring should take account of these dynamics. Where the existing culture is not perceived by staff to value research, the intervention may need to target values, beliefs and leadership prior to (or in parallel with) the other strategies. Attention to an organisation’s history and self-narrative is essential for crafting strategies that will resonate in current circumstances. [293, 349]

Two thirds of the studies used multiple strategies and targeted multiple domains of capacity. This is unsurprising given that supports and constraints in one area of capacity are known to influence capacity in other areas. [245, 253] It is outside the scope of this review to discuss the relative merits of single- versus multi-component interventions—plus others have dealt with this effectively elsewhere [355]—but it does seem that the degree to which intervention strategies are selected, tailored and implemented for local needs and practices is likely to be more important than how many strategies are used, or in what combination. [356] Further, that treating capacity building as a participative endeavour is most likely to generate relevant, locally owned strategies. [246, 354, 357]

The cumulative findings of these studies are a reminder that interventions are complex systems thrust into complex systems. [119] Research utilisation interventions, like other interventions, succeed or fail via their interaction with context: people, places, circumstances and processes will determine what works where, and for whom. Thus the ‘best’ strategies for effecting change are those that are most fit for purpose at the local level. [358] We are warned of the considerable challenges that attempts to build capacity present. For example, that, “…..capacity building is a risky, messy business, with unpredictable and unquantifiable outcomes, uncertain methodologies, contested objectives, many unintended consequences, little credit to its champions and long time lags”. [257:2] Nevertheless, this review shows that there are successes, and informative failures, so we can continue to develop our understanding of how to foster capacity in further interventions.

Implications for future interventions

We note some areas that might be addressed fruitfully in further research use capacity building interventions in policy agencies:

Understanding research use in context. Several studies in our review concluded that they had insufficient understanding of the practices and contexts that were being addressed. This aligns with wider arguments that we continue to have a limited understanding of how policymakers engage with research ideas and
integrate them with other forms of evidence [35], which affects how we conceive of and design interventions, and interpret findings. Designing interventions that are “close to practice” [359] in terms of fit with local research needs and context seems to be essential, but we may also require further investigation of the ‘irrational’ (aka differently rational [219]) and non-linear uses of research that dominate the use of research in policy more generally.

**Researchers’ capacity.** Research-informed policymaking requires that researchers have the skills to produce policy-relevant research, present findings accessibly, and work productively with policymakers; but these skills are often lacking. [219] Six of the reviewed studies attempted to build some aspect of researchers’ capacity in conjunction with that of policymakers [261, 296, 298, 300, 305, 306]; however, a cursory scan of the literature suggests that capacity building for researchers in this field is less developed than for policymakers, with very few intervention trials. The onus remains on policymakers. This appears to be a gap that would benefit from further attention. It would likely require that policymakers are involved in designing the content of such interventions. Many of the mechanisms suggested in our analysis are likely to be relevant.

**Leadership.** Findings reinforced the role of impassioned and strategic leadership as a crucial driver of organisational change, [e.g. 129, 150, 302, 303] including leadership by high profile external experts in the wider policy environment. [294] However, there seemed to be few attempts to target or harness internal leadership in intervention activities. The pivotal role of organisational leaders, champions and opinion leaders in driving change is well established both in practice settings [13, 360-362] and within policy agencies [38, 147, 253, 254]; but we know little about how leadership dynamics within hierarchical and procedure-focused policy agencies function and effect change in relation to research utilisation capacity building. Recent arguments about the strengths of distributed or collective leadership for knowledge mobilisation, including cross-sector partnerships, suggest that our conceptualisation of leadership may need to expand [193, 239, 293]. This area could benefit from further investigation.

**Audit and feedback.** With the exception of Peirson et al. [150], none of the studies reported using organisational-level progress feedback as a strategy, and none used audit and feedback: a process that gathers information about current practice and presents it to participants to facilitate learning, develop goals, create motivation for change, and focus attention on change tasks. [363] Audit and feedback is well-established as a catalyst for professional practice change [363], including the uptake and use of research. [194, 328] There is mixed evidence for its effectiveness, but a recent systematic review found that it generally leads to small yet potentially important improvements in professional practice [364], and it may be more successful than change techniques such as persuasion. [365] It seems a potentially valuable strategy within research utilisation interventions, particularly in the light of systems-influenced implementation frameworks that emphasise the need to establish performance feedback loops in organisational change processes. [12, 223, 293]

**Commissioning research syntheses.** Findings of the two studies that looked at commissioned research syntheses suggest that the value policymakers attribute to syntheses is affected by the commissioning process and/or their involvement in the conduct of the review. [236] [283] A contribution mapping review of 30 studies found that research was most likely to be used when it was initiated and conducted by people who were in a position to use the results in their own work. [366] However, an evaluation of health policymakers’ use of a briefing service provided by academics found that access to the service did not improve the policymakers’ capacity, nor their uptake and use of research. [367] What critical factors are at play in these scenarios? We would benefit from greater understanding of how commissioning models can best support policymakers’ capacity development and use of research, including the contribution that researchers and knowledge brokers can make.
Sustainability. The concept of capacity building is linked to that of sustainability [267], but sustainability itself was seldom mentioned in the reviewed studies. As bureaucracies, policy organisations are characterised by their adherence to protocol, but there appeared to be few attempts to embed strategies within existing work systems (with some notable exceptions e.g. [150, 301]). The need for continuous active participation in knowledge mobilisation practices [267] was evident in few studies. Several tried to embed new knowledge and skills in practice (e.g. via mentored assessment), but this targets individual knowledge rather than organisationally owned processes—an important consideration in organisations known for their high turnover. [246] Sustainability may depend on different mechanisms from those posited here. For example, self-efficacy may be critical for initiating new patterns of behaviour, but have a limited impact on the decision to maintain that behaviour over time. [368] Greater consideration of organisational learning and the use of measures to prevent capacity initiatives from being ‘washed out’ [67] may be required. Longer-term evaluation would help but organisational change, like relationship building, is a lengthy and evolving process, often taking years to reach intended goals. [150, 261]

Underpinning assumptions about research-informed policymaking. Despite the lack of clear theoretical drivers in most studies, the conceptual basis of attempts to address research use in policymaking seems to be maturing: rational linear models of research are being supplanted by ideas from political science, organisational change, systems thinking and other bodies of work that disrupt the evidence-based policy ideal. The field is also making use of opportunities to evaluate capacity building endeavours that are initiated outside of academia; using creative methods to learn from complex real world projects; and refusing to be cowed by the entanglement of change strategies, process indicators and outcomes. However, there is still evidence of “theoretical naivety” as described by Oliver et al. [35]; for example, focusing on research as exemplary evidence rather than on policymakers’ use of diverse information and ideas within which research must function; the belief that a reconfiguration of barriers and enablers to accessing research would lead to greater impact; and a general lack of understanding about policy processes. Oliver and colleagues provide advice about the direction that future research can take to address these gaps. [35]

Strengths and limitations

This paper contributes to our understanding of research utilisation interventions in policy agencies by providing an overview of 22 studies: their change strategies and outcomes, the contextual factors that mediated these effects, and the theoretical perspectives that underpinned them. It also tentatively identifies the mechanisms that can best explain how the intervention strategies achieved their effects, or why they did not. This is an important first step in developing a more theoretically grounded approach to the design and evaluation of such interventions.

The paper may best be described as a realist informed scoping review. [369] Unlike an orthodox realist review it was conducted to inform our own program of research so it was not negotiated with external stakeholders; we did not conduct extensive theory-focused searches; and the analysis was exploratory and inductive rather than an interrogation of program theory. [269] We took an inclusive approach to study design and quality and, given our aim of identifying causal mechanisms, focused on identifying explanations of why an intervention was more or less successful rather than on quantitative findings. [370] The realist perspective contributed importantly to this process by enabling us to identify tentative constructs that may be used to inform the development, implementation and evaluation of subsequent research-to-policy capacity building trials.

The findings are strengthened by independent analyses and critique of draft SCMO configurations, but our limited timeframe prevented the use of strategies that might have strengthened the review’s rigour further such as more thorough searching and contacting the authors of studies for missing information. The distinction between different participant groups and the extent to which the results for different groups could be identified in the findings was not always clear. Including studies in which evaluations focused on
processes and perceptions limits the identification of distal outcomes. The (mostly qualitative) data provided rich clues about contexts and possible mechanisms, but often did not include concrete information about capacity impacts or about any actual use of research in policy processes. Consequently, the findings should be seen as preliminary.

The identification of outcomes is further complicated by the entanglement of intervention strategies and outcomes. For example, improved infrastructure for accessing research, greater advocacy of research by organisational leaders, workforce development and increased interaction between policymakers and researchers can be seen as both intervention inputs and outputs—the means and the ends of capacity building—depending on the focus of the intervention. As such, they tend to be described rather than evaluated. Many of the phenomena being investigated in these studies are complex and evolve over time; a reminder that capacity for using research in policymaking is a work in progress; it will never be fully ‘built’.

Lack of shared evaluation frameworks across the studies means that we were not comparing like with like. A theory-driven approach in which we examined each study in relation to a middle range hypothesis could have produced more focused findings. We took an inductive approach in this first attempt, but believe that subsequent reviews would benefit from a theoretically based investigation. Our review might inform the development of causal hypotheses that further reviews could use as an investigative framework.

Lastly, mechanisms are “squishy” [371]. They change position in SCMO configurations, morphing into contexts and outcomes depending on the focus of the evaluation and level of analysis. [238] They can be differently aggregated, their categorisation is limited by vocabulary and interpretation [371], and their status as causal explanatory devices is uncertain; as Gerring argues, “mechanisms might also be referred to as a theory, theoretical framework, or model, depending on one’s predilection”. [371:1503] Thus the concepts we have called mechanisms, the level of granularity at which they are expressed, and the terms we use to describe them are all uncertain, and many would likely look quite different in the hands of another team. However, we believe that they offer a starting point for further testing and discussion. Typically, middle range theories develop gradually over time based on the accumulation of insights acquired through a series of studies. [372] This review is an early step.

**Conclusion**

This review explores what intervention strategies have been trialled for building capacity to use research in policymaking, and tentatively posits possible mechanisms that might explain how those strategies functioned (or why they did not) in different contexts. The evidence is variable, especially because we included formative and process evaluations that focus more on process effects than measurable outcomes, but our findings suggest that tailored interactive workshops supported by goal-focused mentoring, and genuine collaboration, may be particularly promising strategies. Systems supports (e.g. infrastructure, governance arguments and workforce development) are likely to play a vital role, but it is very hard to disentangle their effects from other intervention strategies and systems flux. Many potential mechanisms were identified, and some contextual factors that appeared to impact the functioning of virtually all intervention strategies. There were some gaps in the reviewed literature that could usefully be addressed in further research.
Part 2.
Framing the intervention and process evaluation
Chapter Five: The SPIRIT study

"... organizational and individual deficits within the public sector have resulted in insufficient capacity to access, interpret and apply different forms of research-based knowledge."

Newman, Cherney & Head, 2017 [149:158]

Chapter Three outlined the value of research-informed policy and the considerable challenges facing interventions that strive to advance this agenda. Twenty-two such interventions were reviewed in Chapter Four which concluded that there is a great deal more work to do globally in developing and implementing context-sensitive strategies within complex organisations and institutional systems, and in determining how such strategies work in different circumstances for different groups of people.

The Australian context is no exception. Despite its relatively small population, “the policy dance” [373] that characterises decision-making in the Australian public service appears to be as complex as other major high-income countries at both the state and national level [127, 226, 373]. Opportunities to improve individual and organisational capacity to access, appraise and use research for policymaking have been identified [149, 226, 374], yet the development and testing of intervention strategies to build capacity in these areas is nascent. SPIRIT was designed to address this need.

SPIRIT provided an excellent vehicle for exploring many of the questions that underpin this thesis. There was opportunity to investigate how intervention strategies functioned in different settings, but also to explore practice norms and values in multiple policy agencies, and to consider the interactions between organisational culture and remits. Overarchingly, it was an opportunity to explore a more fundamental question: What do we need to know about policy practices to help us best support research-informed policymaking?

Supporting Policy In health with Research: an Intervention Trial

In August 2010, the Australian National Health and Medical Research Council (NHMRC) awarded $2.5 million to a newly created Centre of Research Excellence called CIPHER (Centre for Informing Policy in Health with Evidence from Research). Among its aims, CIPHER sought to generate new knowledge by testing an intervention designed to increase the capacity of health policy agencies to use research, and developing techniques for evaluating such an intervention. This study was named SPIRIT - Supporting Policy In health with Research: an Intervention Trial. [227] The multidisciplinary and cross-sector investigator group behind SPIRIT comprised researchers, policymakers and knowledge brokers from Australia (mainly) and the UK.

As shown in Figure 7.2 (on page 101), an ‘action framework’ was developed to articulate and guide the testing of hypotheses about: the catalysts required for research use, the individual and organisational capacities that affect research engagement, and how research use outcomes might be facilitated. [375] The study was designed as a stepped wedge cluster randomised trial. Six health policy agencies that developed and implemented state-wide or national health policies and programs participated. Each agency was randomly allocated to one of three start dates for the year-long intervention. Figure 5.1 shows the timeline of intervention and evaluation activities for each agency. See the study protocol for further details. [227]
Sampling

The agencies that participated in SPIRIT were selected from an initial sample frame of nearly 200 organisations drawn from government websites that listed all New South Wales (NSW) and Australian government health policy agencies. An agency was eligible if: a. its primary work was in health policy development, b. at least 20 staff members were involved in designing, implementing and/or evaluating health policies, and c. the agency was based in Sydney where the SPIRIT administration was located (we were only able to provide intervention activities locally).

Members of the investigator team with policy experience reviewed information from the website of each agency and excluded those without a significant focus on health or on policy design, development or evaluation. Seventy-five agencies were classified as potentially eligible and email and/or phone contact was made to each site to determine the number of staff members who were directly involved in health policy work. The remaining 16 agencies were ranked to determine those with the greatest specific focus on health and the largest numbers of relevant staff. Two investigators visited the CEO or equivalent of the top six ranked agencies and invited them to participate in the SPIRIT trial. All agreed. These six agencies were: a division of the NSW Ministry of Health, and five statutory agencies, four of which had state mandates, and one which worked nationally. Staff in all of the state-based agencies were government employees, while the national agency was technically non-government, but entirely funded by the federal Department of Health.

A written agreement was developed between each agency’s CEO and the chief investigator that detailed: what the agency would receive from SPIRIT, the randomised start dates, and what was expected of agency staff in terms of internal administration and data collection. A member of staff at each agency was nominated by the CEO to act as the liaison person for the study.
SPIRIT intervention design

Figure 9.1. (on page 134) presents a high-level overview of the design, including the principles that guided it, key intervention targets and strategies, and agency-based implementation supports.

The intervention was one year long. During this period, all agencies received six intervention components (Table 5.1). These components included some core deliverables but also content that agencies were invited to customise to address their preferences and priorities. Some components had been derived from an analysis of policy agency capabilities that support research use [374], some were existing strategies and programs used by the Sax Institute\(^\text{11}\) where the research team were based, and others were bespoke for the intervention. Details about the intervention components and their levels of flexibility are included in subsequent chapters, and overview of SPIRIT can be found in the study protocol. [227]

Table 5.1. SPIRIT components and activities, and data collection used in their process evaluation

<table>
<thead>
<tr>
<th>Intervention components</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Audit, feedback and goal setting</td>
<td>• Audit feedback forum</td>
</tr>
<tr>
<td></td>
<td>• Mid-intervention feedback</td>
</tr>
<tr>
<td></td>
<td>• SPIRIT newsletter</td>
</tr>
<tr>
<td>2. Leadership program</td>
<td>• Supporting organisational use of evidence</td>
</tr>
<tr>
<td></td>
<td>• Leading organisational change</td>
</tr>
<tr>
<td>3. Organisational support for research</td>
<td>• Quarterly email endorsement of SPIRIT from agency CEO</td>
</tr>
<tr>
<td></td>
<td>• Access to an online research portal called Web CIPHER</td>
</tr>
<tr>
<td>4. Opportunity to test systems for accessing research and reviews (brokered services)</td>
<td>• Brokered commission: a. a rapid systematic review OR b. an evaluation plan OR c. an analysis of linked data</td>
</tr>
<tr>
<td>5. Research access</td>
<td>• Three occasions of research access from two modes: a. interactive forums with researchers and/or b. an e-bulletin (summary of systematic reviews)</td>
</tr>
<tr>
<td>6. Educational symposia for all policy staff</td>
<td>• Valuing research symposium</td>
</tr>
<tr>
<td></td>
<td>• Two symposia from: a. access to research, b. appraising research, c. evaluation or d. working with researchers</td>
</tr>
</tbody>
</table>

SPIRIT was informed by composite theory including: cognitive behavioural theory, the literature on research utilisation, organisational change, systems thinking and adult learning theories. This theory was articulated in the SPIRIT action framework mentioned above [375] and a list of change principles (Table 7.3 on page 100). Core concepts and definitions were honed through a deliberative process described in Chapter Six.

Outcomes

SPIRIT’s primary outcome was to determine whether the intervention caused an increase in the extent to which research was used in the development of policy and program documents, and the extent to which policymakers undertook four research engagement actions (i.e. accessing research, appraising research, generating new research or analyses, and interacting with researchers). Secondary outcomes using self-reported measures were to determine whether the intervention resulted in an increase in: research engagement actions, the use of research, and the capacity of participating agencies to use research.

\(^{11}\) See https://www.saxinstitute.org.au
Capacity was assessed in terms of the value that policymakers and their agencies placed on the use of research, the confidence of policymakers in undertaking research engagement actions and in using research, and what systems and tools were in place to support research use. [227]

SPIRIT implementation

SPIRIT hinged on an audit and goal-setting process. Eligible staff were asked to complete an online self-report survey prior to the intervention, and a senior member of staff participated in an interview about the systems and structures the agency had in place to support the use of research.

The active phase of SPIRIT began with a feedback session for each agency’s nominated leaders. The session ran for two hours and was facilitated by the chief investigator as an informal discussion. It encouraged agency leaders to draw on the audit data and their organisational knowledge to consider the current strengths of the agency in using research, and opportunities for improvement. A series of goals were agreed by the agency which they were then encouraged to use to guide decisions about which intervention options to select from a menu of potential components.

Workshop components (which accounted for the bulk of the intervention) were delivered by specially commissioned high-profile experts in policy, research or knowledge exchange, as per the requirements of each workshop. The liaison person worked with the SPIRIT knowledge broker and study director to select workshop topics and hone the contents. Agencies received approximately one intervention activity per month and there were no costs to agencies for any aspect of the intervention, other than staff time. Between 32 to 74 staff participated in each site.

Outcome evaluation

During the trial, outcomes data was collected on six occasions at six monthly intervals. Three outcome measures were used each time:

1. SEER (Seeking, Engaging with and Evaluating Research) was an online self-reported survey that sought to ascertain: how eligible policymakers currently use research, their confidence in undertaking tasks associated with using research, how they value using research in their work, and how supportive they feel their workplace is of research use. [376]

2. ORACLe (Organisational Research Access, Culture and Leadership) was a structured interview with a senior member of staff from each agency. ORACLe assessed agency support for research use across seven domains: 1. Processes to encourage or mandate consideration of research in policy development, 2. Tools and programs to assist leaders of the organisation to support research use, 3. Staff access to training in using research, 4. Strategies to help staff to access research, 5. Methods for generating new research, 6. Evaluation of organisational policies and programs, and 7. Strategies to strengthen researcher/policymaker relationships. [377]

3. SAGE (Staff Assessment of enGagement with Evidence) sought to ascertain the extent to which agency staff used research in the development of policy and program documents. At each data point agencies were asked to provide four policy documents that represented their best use of research over the previous six months. The policymaker most involved with each document’s development was invited to be interviewed. SAGE assessed research use across two domains: 1. Research engagement actions (accessing research, appraising research, generating new research or analysis and interacting with researchers), and 2. Research use (instrumental, tactical, conceptual and imposed). An expert panel assigned SAGE scores. [378-380]
The SPIRIT process evaluation

“[Evaluators] are faced with invisible cats wandering within poorly defined boundaries. Our challenge is to herd as many of them as possible into a particular corner of the territory”

Morell, 2010 [381:22]

The role of process evaluation

The task of process evaluation is to capture information that helps explain how an intervention had its effects. According to the UK Medical Research Council, this includes information that “can be used to assess fidelity and quality of implementation, clarify causal mechanisms and identify contextual factors associated with variation in outcomes.”[17]

Process evaluation is increasingly regarded as an essential part of intervention trials, including RCTs [382] and organisational-level interventions [5], but it differs widely from study to study. Like other forms of research, it reflects the researchers’ hypotheses about causality and scientific methods which are, in turn, underpinned by ontological and epistemological beliefs. However, there does appear to an overall trend moving from a positivist to a broadly post-positivist paradigm. The former focuses on quantitative performance data where the intervention is likened to a drug and contextual factors are often regarded as variables to be managed, while the latter tends to take a mixed method approach that frames the intervention as a social process and regards context as an integral part of the inquiry. [383]

This thesis capitalises on the process evaluation I conducted in parallel with the SPIRIT trial12. [384] The investigators wanted to “tell the story of the intervention” which provided me with opportunities to design an in-depth qualitatively-orientated process evaluation which explored the how and why questions that drive my research. This work intersected with two other important roles:

1. SPIRIT knowledge broker. This member of the research team was the ‘face’ of SPIRIT in each agency. She managed local tailoring with liaison people, oversaw the implementation and attended the intervention activities. Together with myself, the knowledge broker had the most intimate understanding of what was happening in each agency and how it was being received.

2. Liaison people. Each agency’s CEO nominated an internal member of staff to coordinate SPIRIT in their agency. Liaison people were a hugely important source of information for me so, where possible, I interviewed them twice and maintained close contact with them throughout the trial. Chapter Eight explores their role in detail.

Data collection

Data collection began prior to the trial, commencing from the point the agencies were first invited to participate, and concluded with a forum looking at post-trial results which agency leaders were invited to attend. Detailed information about data collection in relation to each sub-study is provided in subsequent chapters so here I concentrate on the bigger picture of what I did and why, illustrating the use of “controlled opportunism” where I tried to take advantage of data opportunities in relation to evolving lines of inquiry.

12 As mentioned before, I was not involved in designing the trial or in implementing the intervention
Given the importance of context in my research, Appendix 5 provides an overview of how this was conceptualised.

**Participant interviews: sampling**

Sampling was for maximum variation, purposefully identifying interviewees who had different experiences or views of the intervention trial, and of using research. Four strategies were employed:

1. Following the initial implementation of SEER (an online survey about research attitudes and behaviours that was administered as part of the pre-intervention audit), I invited people to participate in the process evaluation who scored at the highest and lowest ends of the spectrum in questions relating to: how much research they used, how much they valued it, and how confident they were using it. SEER included a disclaimer stating that responses might be used this way.

2. Identifying participants in intervention workshops who expressed divergent views about individual or organisational research use, or aspects of the intervention. Data collection here involved ad hoc conversations or interviews, depending on each person’s availability.

3. Asking interviewees to nominate anyone who could tell me about alternative experiences and views. This is a critical strategy within the constant comparative approach where counter-data is actively sought, and emerging hypotheses are tested against the fullest possible range of perspectives. [386, 387] Also for realist evaluation where understanding variation across program recipients and settings is critical for identifying what works for whom in what contexts. [15, 23]

4. Seeking out people at different levels of the organisation (e.g. junior policy staff, middle managers, members of the executive) and, where it seemed relevant, in different teams or with other attributes. For example, in considering who SPIRIT was working for I realised that relatively new members of staff were describing the value of the intervention differently than their more established colleagues, so I made sure I spoke to some newer staff in each site.

**Participant interviews: design and conduct**

Interview questions were informed by constructs in the literature that were hypothesised to be particularly important (Table 5.2). Appendices 6 and 7 contain the interview questions used at both stages, however these questions evolved as the interviews progressed, particularly in the first phase where a structured exercise was used to explore how research findings fit in with other types of evidence use, and their relative importance. In practice, this exercise led to a stilted conversation that took interviewees out of the particulars of their everyday work and nudged them towards generalised statements about research (e.g. “We all read journal articles here...” and “I use systematic review if they’re relevant”), whereas the nuances of evidence use and their interplay were of greater interest. It transpired that broader questions such as “What forms of information are most useful to you currently?” followed by prompts that dug into their ideas, values and practice examples, were more productive. The interviews were conversational where possible, and structured so there was scope to pursue emergent themes.

Structuring interviews required a balance. As an interviewer I draw on communication skills from my background as a social worker and counsellor (e.g. empathy, congruence, respect and reflecting back [388]) and try to treat interviewees as partners in the inquiry rather than as research subjects. However, as an interviewer conducting process evaluation interviews my objective was to capture explanatory information rather than elicit people’s stories, so I had to maintain control of the interviews and redirect interviewees when they strayed into areas that appeared unrelated to my aims. [48]

---

13 Interviewees were given cards with different categories of potential ‘evidence’ on them, and asked which they used for what purposes, how important they were, and how they interacted, and if any important forms of information were missing. See Appendix 6
This process was strongly influenced by realism and other texts on theory-driven evaluation. Interviewees were guided to think about the conditions that affected how the intervention was perceived and experienced (contexts), what impacts it had – positive, negative and unexpected (process effects and outcomes) and the things that may have caused this (potential mechanisms). They were often asked their views about emerging hypothesis, and I played devil’s advocate when testing the plausibility of potential hypotheses. However, these interviews did not strictly follow Pawson and Tilley [23], and more recently Manzano [48], where the central subject matter of the interview is theory and interviewees are there primarily to confirm or falsify and refine it. In this “teacher-learner cycle” model the researcher teaches the theory to the interviewee who then critiques it, educating the interviewer. [23] The SPIRIT process evaluation was not guiding by an overarching theory so the interview questions covered a range of hypotheses which, naturally, narrowed as the data collection progressed, but remained quite exploratory.

Interviews were recorded and professionally transcribed. Transcriptions were read and corrected, and individual and agency identifiers were removed. Where the process evaluation contributors had access to transcripts they knew the agency code (A1, A2 etc.) and why that interviewee was identified as a useful informant (e.g. their SEER score showed they had high confidence using research, or they criticised aspects of SPIRIT in a workshop, or they had an organisational role that I had not yet learnt about).

Other interviews

Members of the SPIRIT team were interviewed in order to capture information about two important additional events. 1. The conversation between the study’s chief investigator and the CEO (or equivalent) in each agency. In these meetings SPIRIT was explained and the organisation was formally invited to participate. Data of interest were: the CEO’s views about their organisation’s current research capacity needs, goals for participation and any concerns about SPIRIT. And 2. The ‘midway feedback’ where the chief investigator and the knowledge broker met with the executive group in each agency (usually during a management meeting) to provide an update on the latest outcome measures data and discuss any matters arising. Data of interest were: responses to feedback, perceptions of SPIRIT to date and any concerns. These interviews were audio recorded by the knowledge broker and the data synthesised in framework matrices for easy cross-agency comparisons.

Prior to the intervention, I instigated many discussions with the research team about their causal hypotheses and the theory, ideas and experience-based hunches that informed the design and implementation of SPIRIT. These were used to frame the focus of the process evaluation (discussed below), including the participant interview questions. [48]

Fieldwork observations

There were three components to the observations:

1. Monitoring the implementation fidelity and theoretical fidelity of the intervention workshops. This involved identifying and critiquing the ‘essential elements’ of the intervention, and refining them while the trial was underway. Chapter Seven looks at this in detail.

2. Gleaning as much information as possible about people’s experience of the intervention, and about agency culture and practices, from the interactions observed. I developed a framework of questions to

14 The questions were: i. How is policy/program work approached in this agency? ii. What does business-as-usual look like? iii. What information, ideas and process are usual? Which are most important? iv. How is research or research-informed information considered and used in this mix? v. What responses are there to SPIRIT? vi. What else is going on that might affect responses to SPIRIT? and vii. How could research, researchers or research utilisation strategies better meet these policymakers’ needs?
guide these observations and the management of my field notes, but found they were too restrictive and ended up using a more inductive approach where I scribbled notes during and after workshops, and reviewed them for themes. I also audio recorded the workshops and used recordings to go back over especially rich conversations to ensure I had captured the full range of views.

3. Talking informally with participants before and after the workshop. These conversations generally focused on expectations/perceptions of the content and its connection to their work, but often revealed important information about wider issues such as why a division of the organisation had not attended, how the last round of data collection had impacted perceptions of the intervention, or how the agency’s CEO had talked about SPIRIT in a management meeting.

**Considering more extensive fieldwork**

In the early stages of the process evaluation I planned a more ethnographic approach in which I would spend time immersed in each agency (or, perhaps, a selection of them), observing routine practices and trying “to get inside the way they actually put their knowledge to daily use”. [31.6] Intensive fieldwork observations are valuable for understanding work practices [77, 389] and organisational knowledge structures and use [77, 390], including how meaning is constructed in policy processes. [391] Fieldwork observations also help determine how and why interventions work e.g. [e.g. 265, 392], by providing “context evaluation” [67] and exploring interactions and processes in real time from emic and etic perspectives. [393]

My plan for intensive fieldwork observations, however, was met with scepticism by policymaker investigators on SPIRIT. They advised that observations of behind-the-scenes work might be regarded as unacceptably intrusive, particularly in the context of an intervention trial where people would be more inclined to feel they were being assessed. They also questioned what would be achieved given that the much-talked-about ‘policy decision’ is not a discrete event but a messy evolution involving myriad people over uncertain lengths of time.\(^{15}\)

I was reluctant to let this idea go and wanted to see if there were opportunities that could be leveraged with the right approach, so I spent a day in a policy organisation trying to get a feel for how embedded fieldwork could work in this context. Gaining permission from a division in the NSW Ministry of Health that was not involved in the trial, I attended large and small meetings (excluding those that were deemed ‘off-limits’). My request to ‘hang out’ in the main office space, attached to specific members of staff who I could mine for information, was denied. The experience was illuminating, but not as intended. Attendees at the meetings seemed uncomfortable and appeared to be cautious about expressing views, but they also did not ask me questions that might have addressed their concerns (their managers and I had explained my presence, but it was cursory). They seemed keen to minimise contact and rushed off after the meetings so I was unable to talk with anyone one-on-one.

These problems may have decreased if observations were conducted over a longer period, particularly if I was able to establish relationships and demonstrate non-judgemental goodwill (perhaps as a participant observer). But the wider design of SPIRIT and limits on my availability meant that the necessary long phase of relationship-building was not feasible. Further, given the trial context, the underlying concern that I was there to evaluate policy practices rather than understand them may well have endured. As it was, exclusion from ‘sensitive’ areas of work, from people’s routine working space, and from informal conversations meant that little data of interest was likely to be available. Coupled with the evident discomfort it caused, and the potential negative impact on the trial, I reluctantly concluded it was an unproductive strategy within this study.

\(^{15}\) This advice strongly echoes Weiss’ assertion, mentioned in Chapters Three and Six, that most policymakers do not see themselves as ‘decision-makers’ due to their role as cogs in a large bureaucratic machine (Weiss 1980).
Others have encountered discomfort, suspicion and limited access when conducting (or trying to conduct) fieldwork in policy organisations [394-396]. Mosse [397] provides a sobering account of what can happen when participant ethnography is conducted but later rejected by members of the policy organisation. I will come back to this topic in Chapter 10 where I discuss some positive and illuminating examples of intensive fieldwork in policy agencies.

**Participant feedback forms**

These forms were distributed prior to each intervention workshop and completed anonymously immediately afterwards, placed in a folder by the participants. They comprised Yes/No ratings on six statements: 1. The workshop was interesting, 2. The workshop was relevant to my work, 3. The workshop was realistic about the challenges and constraints of our work, 4. The presenter had appropriate knowledge and skills, 5. It is likely that I will use information from this workshop in my work, 6. It is likely that SPIRIT will benefit my agency. Some workshops had additional items, e.g., the forms for audit feedback forums included items about the clarity of the data and participants’ confidence that SPIRIT would be adequately tailored for their agency. All forms contained three open-ended questions: 1. ‘What worked well?’, 2. ‘What could be improved?’, and 3. ‘Any other comments?’. Chapter Nine has more information on the use of these forms and their limitations.

**Other data collection**

To develop a thorough picture of the organisational context in each site, two data sets from the first outcomes measurement point were reviewed. These were: 1. The transcripts from interviews with senior agency staff about organisational level capacity for supporting research. This interview (called ORACLe) addressed concepts such as resourcing, values, leadership and workforce development. And 2. Transcripts from interviews with policymakers who had been involved in the development of a recent policy or program. This interview (called SAGE), focused on specific instances of seeking, appraising and applying research in the process of developing that policy or program, and explored any constraints and enablers experienced by the policymaker. These transcripts enriched my understanding of the practice culture, including how research-engaged the agency was pre-intervention. After this, there was no data sharing between the process evaluation and the outcome evaluation (apart from access to SEER scores as described above).

I also reviewed a range of publicly available organisational documents. These were treated as ‘artefacts’, i.e. explicit and observable aspects of the organisational culture. [258] Artefacts included agency websites, their most recent annual reports, policy documents provided as part of the SAGE interviews described earlier and, where publicly available, their current strategic and operational plans. Agency recruitment was monitored from the point where their participation in SPIRIT was formalised, in order to capture information about initial and changing research-related roles. The information from document reviews was collated in a spreadsheet (in the case of recruitment information it was captured in the running memo for each agency) and integrated into the schematic case studies, combined with information data from SAGE and ORACLe interviews, and the early round of process evaluation interviews.

Written informed consent was obtained from all participant interviewees and workshops attendees (the latter via a combined consent form and sign-in sheet). The workshop sign-in and consent form, the fidelity coding check list and the participant feedback form were trialled during pilot testing of the audit and feedback workshop in a non-participating government agency. All were adapted to be more streamlined following helpful feedback from these policymakers.
A theory-informed approach

Ideally, process evaluation (and evaluation using a multiple case study approach [103]) should be guided by program theory underpinning the intervention design, using constructs or propositions to determine what to look for and how best to capture relevant data [6] [398], including what aspects of context and process are likely to be most important. [15, 23] Thus the unit of analysis is not the intervention nor individuals but hypotheses about the interaction between people, intervention and context. [15, 48, 399-401] In theory-informed evaluation, theories are held lightly, used to raise questions and highlight areas of attention, but not to dictate. The aim is to produce explanations that take account of the multi-factorial nature of causation in complex systems. As Pettigrew warns “Explanations are bound to be holistic and multifaceted. Beware of the myth of the singular theory of social or organisational change”. [402:269]

Table 5.2 shows how the process evaluation attempted to respond to some of the concepts that had both explicitly informed the development of the intervention (ascertained by talking with the intervention designers and implementers) and those that seemed pertinent given the intervention settings (bureaucracies - complex organisations), participants (health policymakers) and intervention goals (increased capacity to use research, and actual use of research, in policy processes). Much of the literature that informed this thinking is outlined in Chapter Two. Note that the review of research utilisation interventions presented in Chapter Four had not yet been conducted. If it had, I would have used the propositions that are tentatively postulated in that review more systematically in the process evaluation.

Data management and analysis

The large amount of diverse data, together with the conceptual complexity of this research, presented some challenges for data management and analysis. I managed this using a combination of strategies:

Running memos

Seven running memos were rigorously maintained throughout the study: one for each agency plus a single cross-agency memo. The agency memos were used to capture and chronologically organise syntheses of descriptive information from multiple sources (using NVivo to link them to the source field notes, transcripts, websites, recordings, and individual memos from interviews, etc), and to develop theories in relation to this data. The cross-agency memo provided a platform for higher level synthesis and analytical development of ideas that were being tested through concurrent data collection. This included my continuous ‘reflective audit trail’. [73] It also provided a comparatively succinct way to share emerging hypotheses with the process evaluation contributors.

Case studies

Case studies were used to collate and synthesise descriptive data pertaining to each agency, and as the foundation of the overarching analysis. They were developed following the process described by Yin [98] (and later depicted in diagram form by Goodrick [103]) using a schematic structure to aid cross-agency comparison (Box 5.1). This structure was developed iteratively using a mixture of a priori constructs and headings identified inductively from early rounds of analysis, and it continued to change throughout the analysis.
Table 5.2. Some propositions identified in the literature and the data collection strategies used to investigate them

<table>
<thead>
<tr>
<th>Tentative propositions</th>
<th>Data collection strategies</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Organisational culture mediates interventions: agencies where research is valued will make better use of intervention resources</td>
<td>Early interviews focused on understand organisational culture and practices from participants’ point of view, e.g. questions about what drives work agendas, how priorities get set, and what influences the use of research. Analysis of ‘baseline’ culture as captured in each agency’s SAGE and ORACLe interviews.</td>
<td>[287, 295, 302]</td>
</tr>
<tr>
<td>2. Organisational role affects what research is valued and how it is used</td>
<td>Participants from across the organisational hierarchy and across different teams within each organisation were sampled, and asked about the interface of role and research use.</td>
<td>[79, 297]</td>
</tr>
<tr>
<td>3. The credibility of workshop presenters affects how policymakers engage with content</td>
<td>Feedback forms included a question about the presenters’ knowledge and skills. Interviews and ad hoc conversations explored what worked and what didn’t work in relation to the presenters. Interactions between presenters and participants in workshops were captured in field notes.</td>
<td>[303]</td>
</tr>
<tr>
<td>4. Involvement in tailoring strategies increases local relevance and applicability, and a sense of ownership</td>
<td>Feedback forms asked about workshops’ relevance and applicability. Discussions with the knowledge broker and liaison people captured information about tailoring processes. Fieldwork and interviews explored whether intervention activities addressed real needs and/or delivered what the agency asked for.</td>
<td>[5, 147, 224]</td>
</tr>
<tr>
<td>5. Policymakers’ existing values, knowledge and experience will shape if/how they use research and how they engage with a research utilisation intervention</td>
<td>Interviews explored experience and confidence, while fieldwork captured immediate responses to workshops and explored participants’ rationales for their responses.</td>
<td>[13, 28, 38, 152]</td>
</tr>
<tr>
<td>6. How participants perceive an intervention will strongly affect how it functions in each site</td>
<td>Interview questions explored perceptions and what lay behind them, and delved into how this impacted on their interactions with SPIRIT and any change effects.</td>
<td>[12, 13, 16]</td>
</tr>
<tr>
<td>7. Interactivity, practice opportunities and other adult learning techniques enhance engagement, learning and self-efficacy</td>
<td>I developed criteria for monitoring levels of interactivity in observations (Chapter Eight) and used fieldwork observations, participant interviews and discussions with workshop presenters and the knowledge broker.</td>
<td>[37, 152, 278]</td>
</tr>
<tr>
<td>8. Managers are key drivers of organisational change; their espousal of a research utilisation intervention and modelling of research engagement will be crucial</td>
<td>Interviews with managers and other staff about the views and role of managers in relation to research use and to SPIRIT. Observations of managers during intervention activities (I knew who they were due to the sign-in and consent process).</td>
<td>[5, 16]</td>
</tr>
<tr>
<td>9. Audit feedback can increase awareness of research capacity needs which may, in turn, motivate managers to support a research utilisation intervention</td>
<td>Observations of audit feedback sessions, and informal conversations and interviews with participants across each agency (including the liaison people) afterwards.</td>
<td>[308, 352]</td>
</tr>
<tr>
<td>10. Archetypal differences between researchers and policymakers may affect research utilisation interventions (e.g. they could be perceived as patronising or lack local applicability)</td>
<td>Interview questions and fieldwork paid close attention to participants’ responses to SPIRIT’s goals, strategies and language, including if perceptions changed over the course of the trial.</td>
<td>[78, 80, 117]</td>
</tr>
</tbody>
</table>
While the case studies were extremely useful for analysis, and for sharing data and refining hypotheses with the process evaluation contributors, they are not provided in publications or in other parts of this thesis because I found it impossible to include a meaningful level of information while also de-identifying them sufficiently. The unique agency characteristics and clearly defined organisational roles necessarily detailed in the case studies would enable anyone who knows or works within NSW Health to identify individuals. Withholding these cases was the only way I could honour my ethical commitment to preserve participant anonymity. This is especially important given the potentially harmful information in the case studies about people’s work practices, plus their frank views about colleagues, their agency’s culture and the wider political system in which they operate. Sadly, the lack of rich case descriptions in this thesis limits the extent to which the story of each agency’s participation in SPIRIT can be conveyed holistically.

Others have encountered similar problems. For example, in their research, McDonnell et al. were faced with a “conscious `trade-off' between making the case studies ‘come to life’ in the final project report and the preservation of anonymity” [403:388], and so felt ethically obliged to report their data thematically. Both process evaluation and case study research are susceptible to this problem, primarily because of the relatively small number of purposefully sampled participants who are usually members of a system which has defined roles and is described in-depth. [403-405]

### Box 5.1. Final schematic structure for the six case studies

1. **The organisational context**
   - a. Agency remit/purpose/priorities
   - b. Agency structure and accountabilities
   - c. People, roles and tasks
   - d. Organisational change history
   - e. Culture and communication
   - f. Agenda-setting and work processes
   - g. Leadership
   - h. Professional development and learning styles
   - i. External relationships
   - j. What would improve work practices?

2. **Relationships to research**
   - a. Organisational support/drivers
   - b. Using research
   - c. Access to research
   - d. Generating research and relationships
   - e. Valuing research
   - f. What would improve research use?

3. **SPIRIT implementation**
   - a. Implementation overview
   - b. Fidelity
   - c. Reach
   - d. Outcome measures
   - e. Process evaluation

4. **Interaction with SPIRIT**
   - a. Relationships with the SPIRIT team
   - b. Leadership involvement
   - c. The Liaison Person
   - d. Participation
   - e. Perceptions/experiences of SPIRIT
   - f. Intervention critiques and improvement advice
   - g. What else was going on?
   - h. Change

### Other analyses

**Framework Analysis** was used to categorise and synthesise interview data in matrices. [406] This enabled the efficient management of large amounts of data (76 transcripts across six agencies at two time points) and ensured that I captured the full range of variation within each category while also having an overview of the whole data set. This method has been criticised for being inflexible and lacking analytic utility; arguably, it is a data management system — a tool for analysis rather than a method — but the development of categories was responsive to the data, allowing room for emergent theories, and it provided a useful platform for further analysis. Framework analysis has the benefit of being transparent and clearly structured which facilitated work with the process evaluation contributors. [406, 407] Category headings used in the
framework matrices can be found in Appendix 8. This approach was used with others to inform the case studies and sat behind the thematic analysis conducted for the study in Chapter Eight, but in reality, it often served more as a ‘table of contents’ for the interview data than a reliable representation of it, so the analysis usually required returning to the transcripts.

Realist analysis was used for the review in Chapter Four and for the overarching analysis of the process evaluation data in Chapter Nine. Data was collected and coded for context+mechanism=outcome configurations in order to identify how intervention strategies were operating under what conditions to generate outcomes for which groups. These analyses are described in the relevant chapters.

Throughout the data collection and analyses I followed the principles of constant comparison which are fundamental in nearly all qualitative work. [386] They include:

- Testing rival explanations, i.e. looking for theories, themes or explanations that counter findings-to-date. [106, 112] Realists also consider and adjudicate between plausible rival theories of causation. [15, 23] This involved emersion in diverse literature and critiquing hypotheses with the process evaluation contributors. I also shared emergent ideas about causality with interviewees and asked them to provide alternative explanations.

- Looking for negative cases. This focuses on seeking data (rather than theories, as above) that may contradict emerging hypotheses. For example, I purposefully identified ‘outliers’ who had different experiences or views of the intervention trial, and of using research. [106, 112]

- Triangulation, i.e. using multiple methods to capture different forms of data and perspectives, and using modes of analysis that enable diverse ideas, themes and lines of enquiry to converge. [98, 404, 408] As Patton explains, ‘... triangulation is based on the premise that no single method ever adequately solves the problem of rival explanations. Because each method reveals different aspects of empirical reality, multiple methods of data collection and analysis provide more grist for the research mill.” [106:1192] Triangulation is also integral to realist analysis [52] and to case study research. [98]

For most of this work I have been blinded to outcomes so that my theorising about possible causal processes, and my interpretation of what I was observing in each site, was not influenced. For example, if I knew that Agency 1 had higher outcome scores than Agency 2 would I view their respective process data differently? It was important to disentangle scores from outcomes in my analysis because I hoped to examine the extent to which my findings foreshadowed the outcomes in each site. For instance, would agencies in which staff had high levels of participation and expressed enthusiasm for the intervention have better outcomes than those with poor attendance and perceptions? Comparing such findings might indicate that some processes contributed to outcomes in important ways, while others did not. This would provide additional information with which to critique the design and implementation of SPIRIT. Now, when I am in the final stages of this thesis, I have been informed of the trial outcomes and I reflect on my findings in light of this knowledge in Chapter Ten.

Ethical approval

Ethical approval for the SPIRIT trial and process evaluation was granted by the University of Western Sydney Human Research Ethics Committee, approval number H8970.
Chapter Six: Defining SPIRIT’s core concepts

“... there is never a single, always-correct definition. When we conduct empirical research, our terms mean only what we say they mean...”

Saylor Academy [409]

Overview

This chapter presents a participant observation study of the process of defining SPIRIT’s core concepts. The chapter that follows asserts that to understand how and why interventions have the effects they do, we need a clear account of what the intervention comprises, how its activities are linked to its outcomes, and how context and program interact. But before we can get to that, bluntly, we need to clarify what we’re talking about. For example, what is this thing called research that we are promoting? Who are the so-called policymakers who could be using research more effectively? What exactly is a policy agency? And how can we define such terms without obscuring the fuzzy boundaries and variation that are so important for understanding real world practices? As Oliver argues, using ‘research’ and ‘policy’ as one-size-fits-all concepts underestimates the variety of activities and outputs involved in each type of process. [35]

SPIRIT was designed and conducted against a backdrop of debate about the core terminology used to describe research utilisation in policy. As Chapter Three illustrates, terms such as evidence, knowledge, policy decision-making and the policy cycle—all of which were crucial to SPIRIT’s program theory [375]—were (and are) in dispute. The investigators who were attempting to use these concepts comprised researchers and policymakers with diverse professional backgrounds and interests who often had different perspectives. An account of how they worked though and reconciled the inevitable tensions sheds light on a neglected area of research and might help others to engage productively in similar processes.

The published version of this manuscript is in Appendix 9.
Manuscript


Abstract

The development of definitions is an integral part of the research process but is often poorly described. This paper details the iterative development of five definitions: Policy, Health policymaker, Health policy agency, Policy documents and Research findings. We describe the challenges of developing definitions in a large multidisciplinary team and the important methodological repercussions. We identify four factors that were most helpful in this process: 1. An emphasis on fit-for-purpose functionality, 2. Consultation with in-context experts, 3. Our willingness to amend terms as well as definitions, and to revisit some methods and goals as a consequence, and 4. Agreement that we would satisfice: accept ‘good enough’ solutions rather than struggle for optimality and consensus.

What’s in a name? That which we call a rose by any other name would smell as sweet

William Shakespeare, *Romeo and Juliet* (c.1597)

In real life, unlike in Shakespeare, the sweetness of the rose depends upon the name it bears. Things are not only what they are. They are, in very important respects, what they seem to be.

Hubert H. Humphrey, Democratic politician and U.S. Vice President (1966)

Introduction

Ludwig Wittgenstein, the Austrian philosopher, illustrates the difficulties of defining terms in his exploration of the word *game*. He argues that definitions of *game* that focus on amusement or competition or rules each neglect import dimensions of the many activities we call games, and thus no single definition can be found. But, he suggests, we probably do not need a definition because we are sufficiently familiar with enough things which *are* games and enough things which *are not* games to recognise the difference between them. [e.g. 410]

This is true for most of us most of the time. We don’t need to define concepts such as friend, beauty, or irony\(^{16}\) because, like the famous US Supreme Court ruling on pornography [411], we know it when we see it. We form views and make decisions based on comparisons and approximations. We cannot know that others perceive concepts in the same way, but we assume that in most circumstances it is close enough. However,

\(^{16}\) Some members of the research team argued that this point is poorly made given the (ironic) success of Alanis Morissette’s 1996 hit single in which she erroneously claims that events such as rain on one’s wedding day, dying soon after winning the lottery, and failure to take good advice are ironic. They are, of course, merely unfortunate.
The development of a definition is an integral part of the research process. Consequently, it is usually iterative and informed by the very process that it is intended to support. Subjects of research become more complex and fragmented as they are better understood, requiring revised distinctions and ever tighter definitions. This increasing precision feeds back into the research process, providing a jumping off point for further investigation, and can also be a valuable output of research. Yet specificity presents particular challenges, particularly in social research. Operational definitions (constructs that define the tangible variables used as indicators in quantitative research) strain to describe socially situated concepts or phenomena. Measuring multidimensional constructs such as disability, anxiety, ethnicity or intelligence is hard; indeed, some argue that it is impossible. [413, 414] Such constructs defy universal application because they are embodied and perceived differently by different people in different contexts. Conceptual definitions (brief descriptions that tell us what a concept means) are more amenable. They refine constructs but can be formulated to cope with complexity and ambiguity, and they can incorporate situational dimensions. But their very precision frequently limits their utility to the conceptual lens and context of the study in which they were developed making genuinely universal definitions something of a holy grail. For example, Locock et al. [360] found that opinion leaders can be important mediators of research-informed organisational change initiatives, but lament the limitations of researching this phenomenon more broadly due to the improbability of capturing the multidimensional and contextual nature of an opinion leader in a single definition.

Definitions also ‘frame’ their subject, asserting how it should be seen. As Hubert Humphrey points out in the quote above, the terminology we use to describe a thing affects our perceptions, including how we attribute value. This has critical implications not only for research, but for ‘real world’ responses too. For example, Huber et al. [415] note that the current WHO definition of health—a ground-breaking contribution to global health made in 1948—now contributes to the medicalisation of society. Laderchi, Saith and Stewert [416] and Green [417] demonstrate that definitions of poverty affect policy agenda-setting, moral attribution and resource allocation. Further, Hodges [418] argues that standardised definitions in ecology can impede progress by preventing important questions from being asked.

Similar concerns have been raised about terminology in the field of research utilisation. The term evidence-based policy has been criticised as naively ignoring socio-political context and the need for negotiated decision-making in a pluralist democracy. [61] Terms that suggest one-way linearity such as knowledge transfer are increasingly supplanted by more fluid and participative terms like knowledge exchange. [215] And, despite definitions of the term knowledge translation that include concepts of exchange and multidimensionality (see, for example, WHO 2005 & NIDDR 2005 in [419]), Greenhalgh and Wieringa [62] argue that the metaphor embodied in the term constrains further study in the field by misrepresenting the socially constructed nature of how health knowledge is produced and used. It seems that terminology and definitions can provide precise in-a-nutshell conceptual syntheses that support critical debate and rigorous investigation, but they can also frame phenomena helpfully. The imperative is to select the best-fit terms and to define them clearly in relation to their context and use.

Some of the more granular research utilisation terms such as policy, policymaker and policy agency are used in a variety of theoretical, empirical and commentary articles without defining them. The danger is that, rather like Humpty Dumpty, their meaning is known only to the people using them. For example, few
articles that explore policymakers’ use of research explain the professional roles their study encompasses: are these policymakers government employees? Political appointees? Elected politicians? Ministerial advisors? A mixture? It is important that we know because the power, constraints and requirements of these roles differ considerably and are likely to affect how research is used. [118] But concepts which are politically volatile and strongly subjective such as policy are particularly resistant to definition. [420, 421] As Smith notes, “the struggle to classify or define policy is itself at the center of political conflict”. [422] Some authors offer procedural definitions, e.g. “A long term, continuously used, standing decision by which more specific proposals are judged for acceptability in terms of means to be employed, ends to be pursued and time frame in which these proposals will have to fit” (Blum in [423]). Some opt for a less tangible approach which takes account of the inaction often associated with policy, e.g. “Something that one group of actors wishes to see carried out by others” [420] and “Whatever governments choose to do or not to do” (Dye 1984 in [424]). While others focus on policy’s diffuse and values-focused characteristics, e.g. “Policy... consists of a web of decisions and actions that allocate... values” (Ham 1993 in [421]).

Definitions of terms such as policymaker will always require some contextual specificity because of the different systems in which policy work is done. Walton & Macagno [425] argue that definitions are not “What a thing absolutely is”, but “What a thing is commonly considered to be, based on evidential considerations pro and contra”. As such, they are works in progress, evolving to reflect our changing knowledge of the world, and adapted for situation-specific use. Ultimately, the value of a definition lies in its functionality for particular purposes [426]: given our goals, how well does the definition capture what we consider this thing to be? In some cases the goal is to develop universal definitions that traverse disciplines and contexts, in others it is to clearly unpack and boundary terms so that they can be investigated.

In this paper we will describe the evolution of five definitions that were developed for the purposes of an intervention study known as SPIRIT (Supporting Policy In health Research: an Intervention Trial). SPIRIT was designed to test strategies for helping government health agencies increase their use of research evidence in policy and program development. Broadly, the study comprises a multifaceted interactive program of tailored education and resource provision, six measurement periods, and a process evaluation. It is being rolled out to six health agencies in Sydney, Australia over two years using a stepped wedge design.

We needed to define five terms for the purposes of this study: Policy, Health policymaker, Health policy agency, Policy documents and Research findings. We intended to use these definitions to: identify bodies of work from which policy/program documents could be selected for review; establish eligibility for an agency-level sampling frame and for individual participation in outcome measures; provide eligibility criteria for selecting documents to be reviewed in interviews; and to provide a frame of reference for participants completing outcome measures.

We have three aims:

1. To illustrate the value of fit-for-purpose conceptual definitions. We do not propose that others will necessarily find our definitions useful—they were developed solely for the purposes of our study—but we do urge research utilisation investigators to explain how they defined their terms in published articles.

2. To provide an account of the challenging process of developing definitions. We hope this will counter depictions of apparently smooth and linear research processes that dominate the literature. [427]

3. To share what we learnt during this process. In particular, the major considerations required to define these concepts, the value of the process of developing definitions, and the factors that facilitated this process.
Methods and results

The definitions were developed to be fit-for-purpose, i.e. to serve specific needs within the study. The purpose of each definition is outlined in Table 6.1, and their development and functionality are explored in turn below.

Table 6.1. Definitions and definitional purposes of terms used in the study

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policy</td>
<td>A formal statement or action plan developed by a government agency or statutory body in response to an identified problem. This includes state-wide or national legislation, policies, programs, directives, protocols, guidelines, and service models.</td>
<td>• Identification of bodies of work from which documents could be selected</td>
</tr>
<tr>
<td>Health policymaker</td>
<td>Someone employed in a policy agency who drafts or writes health policy documents or develops health programs, or who makes or contributes significantly to policy decisions about health services, programs or resourcing.</td>
<td>• Eligibility for agency sampling frame</td>
</tr>
<tr>
<td>Health policy agency</td>
<td>(A body within) a state or federal government department, or a statutory authority, whose focus is to develop policy which has an impact on state-wide or national services and programs intended to improve individual, family or community health.</td>
<td>• Eligibility for agency sampling frame</td>
</tr>
<tr>
<td>Policy documents</td>
<td>A review, report, discussion paper, draft or final policy, formal directive, program plan, strategic plan, ministerial brief, budget bid, service agreement, implementation plan, guideline or protocol with a focus on health service or program design, delivery, evaluation or resourcing.</td>
<td>• Eligibility for selection of documents for targeted interviews</td>
</tr>
<tr>
<td>Research findings</td>
<td>Analyses of quantitative or qualitative data, or theory, found in peer-reviewed papers, technical monographs or books, or in grey literature such as internal studies and evaluations, and reports on authoritative websites.</td>
<td>• Inform the scoring of the outcome measures</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Provide a reference for participants completing outcome measures</td>
</tr>
</tbody>
</table>

The process in all cases was iterative: two members of the team developed initial definitions based on broad searches in academic and government literature, and online reference sources. These were adapted in response to our developing study aims and constraints. Draft versions were reviewed by colleagues working in policy and program development and were revised accordingly except where the feedback did not align with the study design. Definitions were then reviewed by the whole investigator team, revised according to their advice, and signed off.

Technically, that is what happened, but it did not happen smoothly. Despite being aware of the importance of definitions we were slow to engage in detailed dialogue about them. But as soon as we moved into action—testing draft instruments, constructing sample frames, building consultation networks, etc.—the need to be precise about who and what was being sought became pressing. At that point, we found that in some cases no definitions existed, and in other cases they were plentiful but divergent. Some of the richest sources were of uncertain quality, e.g. Wikipedia, blogs, online dictionaries. Then, as we started to consult with colleagues in policy and program work about the face validity of draft definitions, their feedback raised fundamental questions that caused us to critique and amend aspects of the intervention or evaluation design, so we bounced back and forth between the literature, the evolving study design and consultation advice. Our team comprised a large and diverse group of researchers and policymakers located across multiple institutions, cities and countries making it difficult to find time and the appropriate forum to debate disagreements and find a resolution (see also Billings’ [428] account of “angst-ridden” definition development processes within a large research team). We recognised this as an essential and familiar
developmental process, but there were differences of opinion about its duration: at what point would we draw a line in the sand? In fact, we never reached the fabled end point. Definitions that had been ‘signed off’ continued to change in response to shifting perceptions of the study design, goals, contexts and participants right up to the moment we were obliged to give them to our participants. Some are still contested, as we describe below.

**Policy**

*A formal statement or action plan developed by a government agency or statutory body in response to an identified problem. This includes state-wide or national legislation, policies, programs, directives, protocols, guidelines, and service models.*

Our primary purpose in defining *policy* was: (a) to identify streams of work from which policy documents could be identified, and (b) to include the many types of policy and program work conducted by participant agencies because each of them had the potential to influence population health and health services in important ways.

We arrived at our final definition through a process of plunder. We reviewed definitions of *policy* found in dictionaries and in the academic literature, chopped them up and rearranged them to synthesise aspects that aligned with our aims, and dropped or amended aspects that didn’t. For example, consultations with some colleagues who were developing important population level initiatives indicated that they did not regard their work as a form of policy, so we supplemented the definition with examples to capture these: programs, protocols, guidelines and service models. Given that our intention was to identify policy documents, the definition purposefully focused on formal statements or action plans that would have written materials associated with them. The definition does not demand that policy includes action because our intervention focused on getting research into policy and program development processes, but did not extend to implementation or health services practice.

The research team used this definition behind the scenes, often as a way of thinking about health policy agencies because these agencies were defined, in part, by their function and tasks. However, the definition appeared to have minimal utility for study participants. It was included with some study information, especially in relation to the selection of policy documents for review in structured interviews, but a definition was neither asked for nor referred to in any interviews or clarifying conversations. Presumably, policymakers have a better idea than we do of what constitutes policy and do not require a definition. They know it when they see (or do) it.

**Health policymaker**

*Someone employed in a policy agency who drafts or writes health policy documents or develops health programs, or who makes or contributes significantly to policy decisions about health services, programs or resourcing.*

This definition was used to identify individuals within participating agencies. Specifically, to differentiate between staff who would be eligible or ineligible to take part in the outcome measures. Our primary goal was to ensure the intervention reached people who were in the best position to benefit from it— i.e. who
had sufficient opportunity to apply research-related knowledge, skills and resources in their work—so this was where we intended to look for measurable change.

Dictionary definitions of policymaker lacked the specificity we required, e.g. “A person responsible for or involved in formulating policies, especially in politics” (Oxford Dictionaries) or “Someone who sets the plan pursued by a government or business etc” (The Free Dictionary). The academic literature did not present the required detail either since it tends to define the term by who it includes and excludes rather than describing it [e.g. 118, 429, 430]

Early consultation with policy colleagues alerted us to the breadth of roles involved in what we were calling policymaking. They advised us to include staff at different levels within each agency, and to ensure that staff who focused on population-level program development and resourcing were included. This advice, together with our observation that the people we were calling policymakers did not use that term themselves, caused us to reconsider the term policymaker. We liked Dobbins and colleagues’ [431] term decision-maker for its inclusivity, but recognized that this was advisedly broad so as to include health services managers—a group who were excluded from our study. We were also mindful of Weiss’s [432] finding that few policymakers are in the position to make final policy decisions. This is because policy development is a collective incremental process characterised by mutual adjustment, move and counter move, accretion and negotiation, thus policymakers focus on affecting the shape and content of policy discourse rather than deciding on overarching policy. Consequently, given that the term was to be employed primarily for internal use (when we used the definition in agencies it appeared under the heading of eligible participants), and the term policymaker was dominant in the literature, and we continued to use it.

The consultation also raised the murky question of what policymakers actually do—what component tasks make up the policymaking process and how could we identify those that are most amenable to research input? This included tasks such as refining/analysing policy questions, reviewing data, conducting or commissioning research or evaluations, consultation with internal and external stakeholders, working with advisory committees, information synthesis, etc. The catchall phrase...contributes significantly to policy decisions... was added to capture that breadth of activity.

We found this to be a functional definition for our purposes, and feedback from participating agencies confirmed that it provided clear guidance for assessing staff eligibility. For example, it was precise enough to exclude contractors from the study and staff who did not contribute to policy or program development such as those in administration and operations, but inclusive enough to capture many strata of policy and program developers and to allow pilot testing in an agency that funds state health services.

**Health policy agency**

*(A body within) a state or federal government department, or a statutory authority, that focuses on developing policies or programs designed to improve state-wide or population level health.*

We needed to define what a health policy agency was in order to develop a sample frame for organisations that were to be invited to participate in the intervention. Rather like Wittgenstein’s game example above, in the early design stages we knew enough organisations that were *like* our conception of a health policy agency and enough that were *not like* it to categorise organisations as potential participants or not, but we were unable to find any definitions in the literature from which to draw rigorous criteria that could be used to create a sample frame. Consequently, we developed a pragmatic definition that matched study eligibility.
requirements with the goals and in-development methods of the intervention. Our constraints (geographic and financial) were front of mind in this process.

It was agreed that a health policy agency must develop population level health policy or programs as its core business. This excluded health agencies whose primary role is operational, e.g. accreditation, compliance, benefits schemes administration. Given our intervention focused on public health and clinical research, we took a narrow perspective of health which excluded many health-related agencies such as community services. Non-government organisations could have been included in this definition; however, in our study we wanted to test the intervention with larger agencies that had most scope to directly influence population level health—so only state or national level government organisations and bodies with statutory authority were considered.

This definition worked well enough for our purposes, particularly after discussion with agency staff in the early phase of the study led us to refine it to give greater weight to program development, and to remove some unnecessarily granular description.

**Policy documents**

A review, report, discussion paper, draft or final policy, formal directive, program plan, strategic plan, ministerial brief, budget bid, service agreement, implementation plan, guideline or protocol with a focus on health service or program design, delivery, evaluation or resourcing.

The purpose of defining the term *policy document* was to provide inclusion criteria that would enable each participating agency to nominate four eligible documents which best represented their use of research evidence in policy or program agenda-setting, development, implementation or evaluation within a given time period. Key personnel were then interviewed about the process of developing these documents which were considered to be proxies for the policies and program that they related to.

The definition had to be broad enough to allow agencies to select documents that reflected their work, but contained enough to identify the sorts of documents that could usefully be informed by research. The consultation process involved colleagues in policy and program development kindly sending us examples of their documents as well as commenting on the draft definition. With each consultation our overarching definition diminished and our list of examples became longer until, eventually, only a list was left. The range and diversity of authors, content, aims, and intended readers was so great that we were unable to describe these dimensions meaningfully for all eligible agencies in a single brief definition. Happily, we were told that concrete examples were preferable so we honed these and, following a further round of consultation, we dropped *press releases* from the list (they lacked detail and were likely to be less ‘authentic’), and added *budget bids*.

In practice, as in development, this was our most problematic definition. From the start, agencies repeatedly required verbal clarification and expressed confusion. No specific problems with the definition itself were identified\(^\text{17}\), and so no improvements were made, but the concept of a policy document remained ambiguous with a need for locally specific discussion in each site.

\(^{17}\) We did identify problems with the instructions and additional criteria we gave about how to select the documents, but this was not an aspect of the definition and so is not discussed here.
Research findings

Analyses of quantitative or qualitative data, or theory, found in peer reviewed papers, technical monographs or books, or in grey literature such as internal studies and evaluations, and reports on authoritative websites.

The purpose of this definition was to inform the design of the outcome measures (i.e. to identify what sort of research-related information was being used at each measurement point) and to provide a common reference for the SPIRIT team and participants, particularly during interviews that focused on research use.

Definitions of research abound, but they tend to focus on the process of conducting research (systematic gathering and analysis of information to advance understanding) rather than the outputs of the process. Discussions of what constitutes research data provided a useful starting point [e.g. 433] as they drew attention to the important difference between raw and analysed data. Given that most policymakers do not have research qualifications (although a surprising number do), the intervention encouraged the use of analysed data only, and the definition reflected this.

The goals of the intervention were to increase the use of research, but we took a pragmatic view of what that might entail. Much so-called gold standard research is neither relevant nor applicable to policy and program development, or cannot be produced within the necessary timeframes, and grey literature (e.g. government white papers, internal program evaluations) is often particularly helpful. Therefore we sought a definition that was broad enough to encompass this. The definition did not need to address research quality because that was explored conversationally during interviews conducted as part of the study’s outcome measures.

We considered including advice from researchers in this definition, but this was rejected because it was too amorphous a construct and hard to qualify without talking to the researchers themselves. If policymakers sought or used advice from researchers, this was captured in our survey forms and interviews. The definition provided in those instruments includes the addendum: Advice from researchers is considered to be research-informed information, but not research per se.

Few participants asked how we were defining research or referred to it during interviews, but several mentioned their surprise at the inclusivity of the definition given that we were researchers (and thus, by implication, attached to academic hierarchies of evidence). So we concluded that the definition was helpful primarily as an invitation for policy and program developers to talk about the ‘real world’ range of research they used.

The term itself—research findings—was contested. Some members of the team wanted to use the term research as a catch-all for the process of conducting research and for its outputs—the context made it clear which was which. But others argued for greater precision. We also debated using the term evidence. This debate was more challenging since it obliged us to explore our beliefs about the nature and status of human inquiry. Some argued that the term evidence is rhetoric. [34, 61] Others argued that it was common parlance and that authors such as Head [173, 213] and Klein [434] had convincingly expanded the term to include other policy-relevant forms of evidence, but for the sake of clarity we should use the term research evidence or evidence from research. We agreed on the term research findings for use in this study but note that it was used inconsistently (as this very paper illustrates).
Discussion

We managed to arrive at five definitions or example lists that were largely successful for our specific purposes which were to: identify bodies of work from which policy/program documents could be selected for review; establish eligibility for a sampling frame and for participation in outcome measures; provide eligibility criteria for selecting documents for review during interviews; and to provide a frame of reference for participants completing outcome measures. The definitions also played an important role in the design of the intervention and outcome measures. Being explicit about the concepts we were studying obliged us to confront issues that had remained ambiguous in our day-to-day deliberations.

Our draft definition of policy was a pragmatic amalgam of dictionary definitions. Following consultation, we supplemented it with examples of more specific state-wide population health and clinical initiatives. With the benefit of hindsight, a more inclusive response would have been to position program development work more explicitly within the term itself so that we were referring to policy and program development rather than policy. This would also have been a more process-orientated perspective which could have helped us to focus on the range of work practices we wished to influence rather than on outputs which we were already targeting with our definition of policy documents. A similar argument can be applied to our other terms. Each definition expanded to include program development, but the terms themselves did not. For example, having been advised by potential participants who we were calling policymakers that they did not refer to themselves that way, we wrestled with alternatives to the term. We were unable to agree on a single term but, in the light of subsequent changes, policy/program developer may have been more suitable for our purposes. Likewise, the definition of health policy agency worked well, but if we were to use it with participating agencies we would consider revising the term to reflect the importance of program development: health policy/program development agency.

Our definition of policy documents (which we would now call policy or program documents) was designed to help participants identify key documents produced in the course of policy/program work that would be discussed in interviews. Variation between agencies in work processes and document types precluded a description that could be applied to all, therefore the definition became a broad list of examples. However, even this list remained problematic since there was no standard nomenclature: one agency’s discussion paper is another agency’s scoping brief. It is no surprise that, despite our changes to the instructions for applying this definition and the additional inclusion criteria, agencies continued to struggle with it. We recognised that a standalone definition that would apply to all agencies was probably not possible. Given that the process of presenting documents for assessment was an unfamiliar and potentially uncomfortable form of participation, we had an obligation to talk with the people nominating documents about how to apply the definition in their context, and to situate this within a broader conversation about our methods and aims and how the data would be used. Agencies had been given this information in written form, but more nuanced dialogue that took account of local factors and concerns was also important.

Some may disagree with our decision to include grey literature in the definition of research findings, but this reflected the real world use of investigative information within policy, and was acknowledged as such by participants. Including grey literature resulted in in-depth conversations about its use which, in turn, educated the research team about the variety and quality of innovative program designs and fit-for-purpose evaluations conducted by many government health agencies, most of which are never formally published. It was a reminder that much grey literature is grey not because of its merit, but because it is sometimes generated using expeditious methods, or is developed in contexts that do not prioritise academic publishing.

The process of developing definitions—reviewing the literature and other likely sources, synthesising and adapting definitions through the lens of the study methodology and goals, and consultation (testing) with
colleagues in policy/program development—was messy but productive. We identified four factors that contributed to this productivity:

1. An emphasis on functionality: our definitions were intended to be purpose specific so we were able to focus exclusively on our study aims rather than striving for generalisability. We believe that our definitions of policy, health policymakers, and health policy agencies are broadly applicable in different contexts, but they should be reviewed for functionality when used in other studies.

2. Input from a broad range of people with in-context expertise. This helped us to understand the phenomena we were exploring, to refine the parameters of our inquiry and articulate the definitions so they reflected this. The appreciation that program development needed to be recognised explicitly within the field that we were calling policymaking led to changes in the definitions and, hopefully, to more sensitive communication with agencies who were participating in the intervention.

3. We asserted that conceptual definitions can and should be refined in response to data collection. Frustrating as it sometimes felt, the iterative revisions to the definitions obliged us to critique the terms we were attempting to define and, in some cases, to amend particular study aims and methods. We are confident that this resulted in more rigorous definitions, terms and methods. We would have preferred a clearer end-point—greater certainty about having arrived at the ‘right’ terms and definitions—but this will always be something of a judgement call since each application is liable to test them in new ways. Wading around in the marshlands of conceptual thinking and applied problem-solving is in the nature of social research and is critical for sound methodological development, but it must be balanced by the pragmatic need to arrive at an agreed (if imperfect) final decision. We probably underestimated how lengthy this process would be given the complexity of our study, its uniqueness, and the contexts in which it was to be applied. With this in mind, earlier consultation about our definitions and use of a more systematic and transparent revision and consultation process would have been valuable.

4. We agreed early on that the chance of reaching consensus on each of the definitions was, to quote one of the investigators, roughly that of a snowball in hell. The definitions would be intersubjective rather than objective, that is, they would reflect a variety of broadly agreed (or at least explicitly recognised) perspectives rather than providing a factual description. [435] The concept of deliberative dialogue was used where the goal is to explore different perspectives and to reach a shared understanding of the topic that can feed into solutions. [436] This allowed us to move forward with some terms and definitions that were not favoured by every team member. In fact, the debate about their merits and implications continues.

What is in a name? Returning to our opening quotes, we would argue that Hubert’s advice is more persuasive than Romeo’s declaration. The name that we give a thing frames how we see it. Defining a thing makes this frame more transparent and, certainly in our case, can lead to helpful critical consideration of the name itself. Although it is valuable, unpacking and delineating complex social phenomena is a challenging task, even within the confines of a relatively focused study. We take our hats off to those who tackle the fraught goal of developing universal definitions.

Conclusion

This paper explores the development of fit-for-purpose conceptual definitions of five terms: policy, health policymaker, health policy agency, policy documents and research findings. It illustrates the role of definitions as one of the building blocks in study design and provides an account of the process of developing them—reviewing the literature and other likely sources, synthesising and adapting definitions through the lens of the study methodology and goals, and testing draft definitions with colleagues in policy and program...
development. It also describes our learning; in particular, our increasing appreciation of the value of the process itself, and the four factors that were most helpful: 1. An emphasis on fit-for-purpose functionality, 2. Consultation with in-context experts, 3. Our willingness to amend terms and definitions during their early stages of use and to refine some methods and goals as a consequence, and 4. Agreement that the team would ‘satisfice’: accept ‘good enough’ solutions rather than struggle for optimality and consensus.

The development of definitions was a challenging, messy and invaluable process. It allowed us to communicate more effectively with study participants, but it also obliged us to recognise our hazy understanding of key concepts and to hone and articulate them, to test our assumptions about shared meanings, and to wrestle with some fundamental questions about the nature of our inquiry. All of which, we believe, improved the overall rigour of the study. We ask others in the field of research utilisation to define their terms when publishing.
Part 3.
What happened, and why?
Chapter Seven: Flexibility and fidelity

“... an understanding of the causal assumptions underpinning the intervention and use of evaluation to understand how interventions work in practice are vital in building an evidence base that informs policy and practice.”

Moore et al. (2015) [17]

Overview

This chapter addresses some aspects of the ‘what’ and ‘how’ questions posed by the thesis title. It tackles the issue of how to conduct a theory-informed fidelity assessment for locally tailored interventions in complex settings. The focus is on identifying and evaluating the ‘essential elements’ of an intervention rather than concentrating on techniques or activities that may look like the planned intervention, but which may have no capacity to effect change. The goal was to combine a fidelity assessment approach (which, traditionally, tends to favour standardisation and reproducibility) with a contextually sensitive and theory-driven approach that would allow us to refine our understanding of what actually needed to be delivered.

The relationship between ‘essential elements’ of an intervention and causal mechanisms

Given that the following paper focuses on the task of identifying essential elements, it may be useful to first clarify the relationship between essential elements (a concept from fidelity assessment) and mechanisms (a realist concept) so they are not conflated. While these concepts stem from different schools of intervention and evaluation, they share a common philosophical goal: to move our attention away from what Dixon-Woods et al. [2] call the “superficial outer appearances” of an intervention and focus instead on what actually causes change. Both remind us that ‘what works’ in a social intervention is not the number of workshops, length of activities, etc. but something more mysterious. However, essential elements and mechanisms have quite different roles in process evaluation.

Essential elements are aspects of social interventions that must be implemented if the intervention is to be effective. As such, they are largely observable and deliverable techniques, albeit pared down to their most specific functions. Mechanisms operate below this observable level of techniques, at the level of participants’ reasoning. As Weiss explains, a “mechanism of change is not the program service per se but the response that the activities generate”. Thus, a mechanism cannot be delivered, only activated. Further, it may be activated by one or more different techniques, depending on the context. So essential elements are considered to be essential for triggering mechanisms, but they are not mechanisms themselves (Table 7.1).

However, from a realist perspective, the same activity or technique may trigger very different responses, depending on the context. Taken to its extreme, this argument means that no intervention activity can ever be deemed essential – they are all contingent. This would require that evaluation ignores what is delivered and focuses entirely on whether and how the intervention’s hypothesised mechanisms were triggered. Yet

---

18 Realists state that mechanisms operate at a different level of reality to the empirical level and so are not visible: we only know a mechanism exists because it results in observable and undeniably real phenomena. Some find unobservability “repugnant”, but the explanatory properties of mechanisms are nevertheless indispensable for advancing science. Psillos, 2009.
we need to identify essential elements in order to plan, describe, implement, monitor and evaluate the intervention with precision, building sound empirical information about what is most likely to work in given circumstances. Even in the most flexible of interventions we must identify and assess strategies that are hypothesised to activate mechanisms in that context. It seems that if we conceptualise essential elements as conditional, just as mechanisms are conceptualised, they can cohabit and serve their equally important functions.

Table 7.1. Some key differences between essential elements and causal mechanisms

<table>
<thead>
<tr>
<th>Essential element</th>
<th>Causal mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>This is an activity that has to happen in a specific intervention if desirable change mechanisms are to be activated</td>
<td>This is a response generated by interactions between intervention activities, participants and their contextual circumstances</td>
</tr>
<tr>
<td>Can be delivered as an intervention strategy, but is often be seen as a nuance of implementation</td>
<td>Cannot be delivered. Rather, it is activated by the interaction described above</td>
</tr>
<tr>
<td>Concrete actions or events that are mostly observable</td>
<td>Abstract processes that are seldom observable so must be inferred from and tested against empirical data</td>
</tr>
<tr>
<td>Should be the delivered in all intervention sites where fidelity is considered important, but may require adaptations to accommodate contextual differences</td>
<td>Will generate different outcomes for different groups in different intervention sites</td>
</tr>
<tr>
<td>Are important in intervention design, and for implementation planning and fidelity assessment</td>
<td>Are important for all forms of theory-informed evaluation</td>
</tr>
</tbody>
</table>

Figure 7.1 shows how these relationships can be conceptualised within a process evaluation. The figure depicts a hypothetical intervention that is being delivered in a complex setting. Its essential elements operate within the wider suite of intervention activities. Here they are represented like the yolk of the intervention egg, echoing Greenhalgh et al.’s description of organisational interventions as having a “hard core” of irreducible elements and a “soft periphery” that can be adapted for better local fit. [12] As suggested by the red outline, these essential elements are required to activate the mechanisms which will, in turn, generate (or act as) process effects and cause the intervention’s outcomes. This figure shows that the interactions brought about by the intervention strategies in this context have activated three mechanisms. Mechanism 1 resulted in a desired outcome. Mechanism 2 was activated but the process effects meant it was unsuccessful in terms of outcomes. For example, if the mechanism was relevance it may have been activated by consultative tailoring of workshop content and promotion that framed the workshop carefully in relation to current agency priorities, generating interest among targeted policy staff who can see potential value in attending the workshop (a positive process effect). But the workshop presenter may have misinterpreted the brief, or pitched content at the wrong level, or organisational priorities may have shifted suddenly in response to a heavily publicised health crisis, reducing participants’ perception of relevance, and preventing their investment in the ideas and opportunities necessary for desired change. The figure also shows that mechanism 3 was activated so weakly in this context that it simply petered out, and that no other mechanisms were activated.

---

19 If they are contingent, essential elements probably need to be renamed! The term ‘active ingredient’ may serve our purposes better.
In places, the following paper blurs the line between essential elements and mechanisms because some of the putative essential elements are, in fact, mechanisms. For example, one of the essential elements is expressed as: The session content was relevant to the agency's work (Table 7.6). Relevance cannot be delivered and observed, as a mechanism it must be inferred from accounts of participants' subjective reasoning. This is recognised in the data collection strategy for this item—feedback forms, interviews and ad hoc conversations rather than direct observation. Consequently, the lack of clearly defined mechanisms in the paper does not raise questions about its internal validity; I am merely pointing out that, in the context of a thesis with realist tendencies, there is an inconsistency in terminology that should be recognised.

Note that this figure erroneously suggests a linear progression of concepts that do not interact. This is an artefact of schematic representation and is NOT intended to suggest that these complex social and psychological processes have clear boundaries or function in tidy ways. The figure echoes Figure 4.1 which depicts the relationship between mechanisms, process effects and outcomes at different levels of analysis. The concepts from this figure are included here, with the addition of the intervention's essential elements.

This is not uncommon in the implementation literature that attends closely to complexity but does not take an explicitly realist stance. See, for example, work by Penny Hawe and colleagues in the thesis references.
Manuscript


Abstract

**Background**
In this paper we identify and respond to the fidelity assessment challenges posed by novel contextualised interventions (i.e. interventions that are informed by composite social and psychological theories and which incorporate standardised and flexible components in order to maximise effectiveness in complex settings). We: (a) describe the difficulties of, and propose a method for, identifying the essential elements of a contextualised intervention; (b) provide a worked example of an approach for critiquing the validity of putative essential elements; and (c) demonstrate how essential elements can be refined during a trial without compromising the fidelity assessment.

**Methods**
We used an exploratory test-and-refine process, drawing on empirical evidence from the process evaluation of Supporting Policy In health with Research: an Intervention Trial (SPIRIT). Mixed methods data was triangulated to identify, critique and revise how the intervention’s essential elements should be articulated and scored.

**Results**
Over 50 provisional elements were refined to a final list of 20, and the scoring rationalised. Six (often overlapping) challenges to the validity of the essential elements were identified. They were: 1. Redundant - the element was not essential. 2. Poorly articulated - unclear, too specific or not specific enough. 3. Infeasible - it was not possible to implement the essential element as intended. 4. Ineffective - the element did not effectively deliver the change principles. 5. Paradoxical - counteracting vital goals or change principles. Or 6. Absent or suboptimal – additional or more effective ways of operationalising the theory were identified. We also identified potentially valuable ‘prohibited’ elements that could be used to help reduce threats to validity.

**Conclusions**
We devised a method for critiquing the construct validity of our intervention’s essential elements, and modifying how they were articulated and measured, while simultaneously using them as fidelity indicators. This process could be used or adapted for other contextualised interventions, taking evaluators closer to making theoretically and contextually sensitive decisions upon which to base fidelity assessments.

---

22 The published version of this manuscript is in Appendix 10.
Background

The process evaluation literature frequently characterises interventions as a ‘black box’ meaning that little is known about how they function, including the hypotheses that underpin their design. [6, 8, 438] Process evaluation shines a light in this box by investigating ‘how and why’ questions about the intervention’s implementation, change mechanisms and contextual interactions. [17]

Fidelity assessment is a fundamental part of process evaluation. Its purpose is to ascertain “the degree to which an intervention or procedure is delivered as intended”. [439:407] This is achieved by operationalising the intervention theory and monitoring the consistency and congruence with which it is implemented. [440-443] In order to determine if the delivery was “as intended” two areas of assessment should be considered: implementation fidelity and theoretical fidelity. Implementation fidelity tells us to what extent the intervention-as-delivered matched the intervention-as-planned. The assessment focuses on measurable or codifiable dimensions such as how intervention providers were recruited and trained, what proportions of targeted people were reached, the amount of exposure participants had to intervention activities (intervention intensity), and the consistency with which the intervention components were delivered in each setting. [444] This is a comparative enquiry that identifies variation between desired and actual activities, between participant sites, and over the duration of the intervention. Implementation fidelity assessment is vital for understanding the intervention’s variation [443, 445], determining its feasibility [440, 446] and determining whether an ineffective intervention was due to poor implementation or flawed design. [438, 446-449]

Theoretical fidelity tells us the extent to which the intervention-as-delivered was congruent with the intervention theory (the logic and hypotheses that underpin the intervention design [437, 450, 451]). This intervention theory is operationalised in the form of ‘essential elements’: manifestations of the theory—the ‘active ingredients’—which must be implemented if the intervention is to be effective. [8, 440] The assessment uses the intervention’s essential elements as indicators for a formative enquiry that makes judgements about the validity of the intervention design in practice. This helps us determine how the intervention worked or why it did not. [450-452] As the new UK Medical Research Council guidance for process evaluation states,

> It may never be possible to fully understand how variations in delivery affect outcomes, given that adaptations do not occur at random, and will be confounded by factors promoting or inhibiting intervention effects. A strong understanding of the theory of the intervention is a prerequisite for meaningful assessment of implementation, focused not just on the mechanics of delivery, but whether intervention remained consistent with its underlying theory. [453:41]

Ensuring theoretical fidelity is vital for assessing the program theory [448], predicting outcomes [443, 454, 455], translating and adapting interventions for other contexts [446, 452, 456], further developing the intervention’s evidence-base [443, 457], and enabling ‘streamlining’ that may reduce burden and cost. [440, 458] In trials of complex interventions, fidelity assessment supports interpretation of intervention outcomes ensuring that observed effects (or lack thereof) can be linked to implementation of the intervention. More positive outcomes have been observed when interventions are delivered with high implementation and theoretical fidelity [443, 446, 451], including in flexible interventions providing that adaptations are locally and culturally appropriate and are congruent with the program theory. [19, 445, 459-461]

The concept of assessing fidelity as part of intervention evaluation originates from psychotherapeutic programs. The aim of fidelity assessment in this context is to ensure prescribed treatments are delivered with minimal variation [449, 455] and adhere to the behaviour change theory that informed their design.
This approach has proliferated within implementation science and is now used for a range of interventions designed to change professional practice in health care. There is increasing formalisation of the theory that underpins these interventions and their essential elements, leading to testable theoretical frameworks and taxonomies of standardised techniques that support replicability and evidence synthesis across studies [e.g. 462, 463].

However, this approach cannot be used for all intervention trials. Indeed, its proponents do not suggest that methods designed to assess the fidelity of “clinical actions performed by healthcare workers in the process of delivering healthcare” [463] should necessarily be more widely applied. [464] Two aspects in particular pose problems for translation: i. the focus on individual behavioural change, and ii. the specificity with which the theory is operationalised. The former is problematic because the best-developed methods of fidelity assessment identify essential elements from a taxonomy of techniques derived from individual behaviour-change theory. [462, 465] No equivalent exists for interventions informed by broader social science theories that target complex interactive, organisational and system level properties. [67, 444, 466] The latter is problematic because it is too restrictive for assessing the fidelity of flexible interventions designed to allow local adaptation in order to increase their relevance and applicability. [467-469] Nor does it capture how interventions respond reflexively to unique characteristics and unpredictable reactions in their settings. [470] This fidelity/adaptation dilemma [456], is particularly pertinent for interventions based on composite theory that are designed for dynamic real world systems in which it is necessary to balance standardisation of both form and content with responsivity to context. Indeed, resolving the fidelity/adaptation dilemma in these contextualised interventions is one of the most important challenges for evaluation. [471] (For clarity, we use the term contextualised intervention rather than complex intervention in this paper as complex interventions are most commonly defined in relation to structural design rather than their theoretical or contextual characteristics [187]).

A growing body of literature documenting the evaluation of contextualised large scale interventions attempts to tackle the challenges of composite theory, flexibility and responsibility to context. These interventions include those informed by ecological, complexity, empowerment and realist perspectives, and those tailored by local providers or developed participatively [e.g. 68, 167, 265, 467, 472-476]. However, while many studies link their intervention’s essential elements to theory, they seldom report sufficient detail for others to see how that theory was translated into specific intervention techniques (rather than other techniques or variants that might be equally well supported by the theory). Moreover, some assume prior knowledge of the form that the intervention and its underlying theory will ultimately take, failing to acknowledge that an intervention’s so-called essential elements may function as conditional elements: contingent on the interaction between intervention techniques, heterogeneous participants and contextual characteristics. [23, 172, 477, 478] Consequently, the intervention designers may be obliged to make countless incremental adjustments to the techniques and the theory that underpins them while the trial is in progress, thus, “By the end of the program, the designers’ operating theory may look quite different from the theory with which they started”. [2] Intervention studies targeted at community populations such as cultural groups often highlight the contingent validity of program theory and why it should be critiqued, (re)operationalised, and potentially rejected, depending on local needs and conditions [e.g. 19, 476, 479], but this is often lacking in organisational level studies. [478] So few trials conducted in policy organisations have been reported that, currently, our knowledge of how intervention strategies may interact with variations in these environments is little more than speculative.

Despite widespread agreement that all intervention trials should document the extent to which their essential elements were delivered [440, 446, 468], no universal methodology exists for identifying or measuring essential elements [442-444, 480] and, for interventions with composite theory, there is sparse guidance for ensuring putative essential elements are valid indicators of the underpinning theory. [443, 454, 470, 480] So how should we determine which elements of an intervention are genuinely essential and which
can be adapted without impairing effectiveness? Calls for greater attention to these questions are widespread, coming from multiple sectors in health [382, 439, 440, 447, 450, 457, 470, 481], education [452, 480, 482], and community development. [18, 445, 454, 467]

How are essential elements identified?

When based on previous studies, intervention designers can identify essential elements from analysis of earlier interventions or operationalise them using exemplary models that have established effectiveness. [443, 444, 446] Theoretically informed standardised behaviour change techniques are in development, but these are currently limited to interventions founded on psychological theories. [463] When designing and evaluating novel contextualised interventions designers can either articulate the essential elements themselves or consult with expert colleagues. [442, 443, 452, 481] Many evaluations tackle this post hoc, piecing together the essential elements via discussion with the designers and/or by reviewing intervention materials. [446, 452, 480]

The design of interventions in trials is often founded on an amalgam of hypotheses that attempt to take account of inter-related theoretical, contextual and pragmatic factors. These include: formal and substantive theories; hunches based on professional experience; and considerations such as study resources, demands on participants, existing practice and infrastructure constraints. The intervention’s essential elements are representations of these composite working hypotheses. [480] Thus essential elements are not extant change agents waiting to be discovered; rather, they are ways of putting working theories into practice in particular circumstances, chosen as the ‘best bet’ from many potential candidates. [441] It is not surprising, therefore, that newly developed essential elements for all types of intervention need to be assessed in-situ to determine the extent to which they capture and truly deliver the intervention theory in the context of messy real world delivery. [450]

How specific should essential elements be?

The degree with which essential elements are specified must align with the level of flexibility in the intervention design. Minimally specified essential elements are appropriate for highly flexible interventions because they can be interpreted for different contexts. [55, 66, 466] These essential elements tend to be expressed as principles, goals or functions (rather than specific techniques or formats) as these provide scope for diverse implementation strategies. Fidelity rests on the extent to which the resulting strategies align with the principles, goals and/or functions (see [18] for examples). [67, 483] Equal emphasis should be placed on how discretionary elements were tailored, and with what process effects. [18, 67]

Where the intervention combines standardised and flexible components an appropriate balance must be found. Essential elements that are too tightly specified oblige providers to adhere to prescriptive scripts and techniques which may be suboptimal or entirely inappropriate in different contexts and circumstances. [19, 467, 483] Whereas minimally specified essential elements may not provide sufficient concrete guidance for developing or monitoring the core intervention activities. [455] The specificity of essential elements is critical for defining what the intervention is and what it is not, including which elements are genuinely essential and which can be adapted. [447, 480] To date, the literature does not provide the detail needed to identify, or determine the specificity of, essential elements for contextualised interventions.

Aims

In this paper, we identify and respond to the challenges of fidelity assessment in contextualised interventions using the SPIRIT study as an example. SPIRIT (Supporting Policy In health with Research: an Intervention Trial) is testing the effects of a suite of strategies designed to increase the capacity of health
policy agencies to use research. SPIRIT recognises that policymaking is a messy subjective social process that takes place in complex open systems with myriad influences. [375] How research is used in policymaking is not fully understood [35], but it appears that different structures, pressures, relationships, values and events interact to shape its relevance, applicability and use, and that this flux cannot be controlled during interventions. [35, 133, 456, 472] Consequently, SPIRIT draws on diverse theories from social and political science, targets individual and system level capacities, and, as Table 7.2 shows, attempts to balance standardisation with responsivity to context in its implementation and evaluation.

Specifically we: (a) describe the challenges of, and propose a method for, identifying the essential elements of a contextualised intervention (a semi-flexible, theoretically eclectic intervention designed for complex settings); (b) provide a worked example of an approach for critiquing the validity of putative essential elements; and (c) demonstrate how essential elements can be refined during a trial without compromising the fidelity assessment. We consider how this approach might complement current methods for identifying essential elements.
### Table 7.2. The degree of flexibility in SPIRIT intervention components and subcomponents

<table>
<thead>
<tr>
<th>Intervention components (fixed)</th>
<th>Subcomponents</th>
<th>Targeted participants</th>
<th>Degree of flexibility in form and content (none, limited, partial or extensive) *</th>
</tr>
</thead>
</table>
| 3. Audit, feedback and goal setting | a. Feedback forum  
b. Intervention selection  
c. Identification of other strategies  
d. Mid-intervention feedback  
e. SPIRIT newsletter | Senior leaders and other key managers, as determined by each agency | Partial: Tailored presentation based on agency’s audit data. Informal discussion shaped by participants’ interests. |
| 4. Leadership program | a. Supporting organisational use of evidence  
b. Leading organisational change | Senior leaders and other managers depending on size and configuration of agency | Partial: Standardised presentation determined in consultation with providers, but with agency-specific case examples. Discussion shaped by participants’ interests. |
| 5. Organisational support for research | a. Quarterly email endorsement of SPIRIT from agency’s CEO  
b. Access to WebCIPHER (an interactive research portal)  
c. Resources for improving the agency’s use of research | All agency staff involved in policy/program work | Partial: Proforma text that CEOs can adapt |
| 6. Opportunity to test systems for accessing research and reviews (brokered services) | a. Brokered commission of:  
  - a rapid systematic review OR  
  - an evaluation plan OR  
  - an analysis of linked data | Primary: Agency-selected staff who would benefit from experience commissioning a service. Secondary: all staff working in the topic area addressed by the commissioned product | Extensive: Standard brokerage processes are used but agencies choose the product, and specify the topic and how it should be approached to best meet their needs. |
| 7. Research access | Three occasions of research access from two modes:  
a. Interactive forums with researchers AND/OR  
b. Summary of systematic reviews | All policy/program staff working in the topic area covered by the forum | Extensive: Agencies choose the topic. They can nominate a particular provider and negotiate the form of the session. Providers shape the delivery details. |
| 8. Educational symposia for staff | a. Valuing research symposium  
b. Two symposia from:  
  - Access to research  
  - Appraising research  
  - Evaluation  
  - Working with researchers | All agency staff involved in policy/program work | Limited: Agencies can nominate case examples |
|                                |                | All policy/program staff who might benefit from enhanced skills in the techniques covered | Partial: Agencies select symposia topics from the menu. They can tailor the focus, nominate case examples and providers, and negotiate the form of the session. Providers shape the delivery details. |

* In all cases agencies had the scope to negotiate session dates, times, duration (between 1-2 hours) and attendance.
**Context for this study: SPIRIT**

Our fidelity assessment was developed and conducted as part of the process evaluation of SPIRIT (Supporting Policy In health with Research: an Intervention Trial). In this trial, six health policy and program agencies based in Sydney, Australia, participated in an intervention designed to increase the capacity of policymakers and program developers to use research in their work. SPIRIT was informed by cognitive behavioural theory, systems thinking, the literature on research utilisation, organisational change and adult learning theories. These were articulated in the form of the SPIRIT Action Framework (Figure 7.2.) and a list of change principles (Table 7.3.) which, in turn, guided the intervention design, and the goals and strategies of individual activities. [227, 375]

**Figure 7.2. The SPIRIT action framework**

The intervention comprised multiple components: (i) audit, feedback and goal setting; (ii) a leadership program; (iii) organisational support tools; (iv) the opportunity to test systems for accessing research; (v) research access workshops; and (vi) educational symposia. These components had varying degrees of flexibility as outlined in Table 7.2. Agency staff received approximately 11 face-to-face sessions over the 12-month intervention period, combined with periodic feedback and ongoing access to resources. Proximal and distal outcomes included: 1. Organisational capacity to use research (staff knowledge, skills and perceptions of the value of research, and organisational support for the use of research as demonstrated through leadership support, policies, tools and systems), 2. Research engagement (accessing, appraising and generating research, and interacting with researchers), and 3. Research use in policy or program work (demonstrated through the assessment of nominated policy documents). Agencies could prioritise outcomes they wished to improve by tailoring the intervention, e.g. to target particular knowledge or skills.

High profile policy and research experts were recruited to deliver the face-to-face intervention sessions. The outcome measures comprised an online survey and two structured interviews. Further details are provided in the study protocol. [227]

### Table 7.3. SPIRIT change principles

| Systems framework | • Uses a multi-component approach  
|                   | • Maximises interaction between the different components of the intervention  
|                   | • Addresses systems, operations, structures and relations  
|                   | • Is flexible in meeting the needs of different agencies  |
| Engagement and ownership | • Engages agencies in owning and driving the program  
|                   | • Is tailored to focus on the agency’s priorities  |
| Goal setting and feedback | • Provides feedback about current practice  
|                   | • Provides a clear rationale for change  
|                   | • Develops agreement about concrete and specific change goals  
|                   | • Monitors and provides feedback about change during the intervention program  |
| Interactive skill development | • Provides self-education opportunities and access to resources  
|                   | • Recognises the expertise of participants  
|                   | • Is interactive with a focus on shared reflection and problem solving  
|                   | • Provides opportunity for rehearsal and practice  |
| Leadership, roles and relationships | • Uses champions to model and promote the use of evidence from research (including both internal and external champions)  
|                   | • Uses credible, dynamic experts as presenters  |

### The challenges

Several characteristics of SPIRIT presented challenges for fidelity assessment. Addressing these challenges drove the methods we used:

1. **Composite theory.** The intervention was built on cross-disciplinary composite theory that had not been operationalised in previous trials. This theory was articulated in the SPIRIT Action Framework and change principles (Figure 7.2 and Table 7.3) but these did not identify which intervention elements should be used as fidelity indicators, nor the level of specificity with which they should be operationalised.

   The manner in which the essential elements should be articulated was complicated by the paradigmatic tensions and different fidelity traditions in the composite theory. For example, cognitive behavioural theories lean towards positivism and experimental intervention approaches, and fall within the
standardised approach to fidelity assessment outlined at the beginning of this paper in which essential elements are tightly specified. Systems thinking, on the other hand, proposes a complexity-orientated ecological worldview in which interventions are loosely specified for local adaptation and essential elements are articulated as principles rather than concrete techniques. SPIRIT, like many contemporary interventions, was occupying a middle ground.

2. **Flexibility.** The expression of the essential elements needed to accommodate three levels of flexibility: (a) agencies were able to select different session options from a menu of components, (b) they could tailor the topics and goals of these options to address local priorities, and (c) expert providers determined the detail of delivery (see Table 7.2). We could not foresee how these decisions would shape the content and form of the intervention. Given that meaningful comparison of the extent to which essential elements were delivered required that they be equally applicable across all intervention sites, our fidelity criteria had to cover both standardised and locally adapted intervention components and reconcile potentially disparate adaptions.

3. **Responsivity to context.** The implementation plan was not fully developed when the trial commenced and was going to incorporate a degree of responsivity to shifting agency priorities, so we needed capacity to adjust our fidelity criteria and data collection methods as the need arose. The complexity of the intervention and of the participating organisations precluded any confident prediction about the essential elements’ validity (would they accurately reflect the intervention theory? would they turn out to be essential?) or even their feasibility (could they be implemented as planned?).

**Methods**

As a result of these uncertainties we were unable to predetermine the content, scope and specificity of the essential elements. Consequently, we judged it necessary to identify provisional essential elements and observe them in the field, using empirical evidence from the process evaluation to revise them as required. Our goal was to critique the construct validity of the essential elements [443] and modify them while simultaneously using them as reliable fidelity indicators.

The mixed-method process evaluation focused on three domains: (a) how the intervention was implemented (fidelity assessment), (b) how people participated in and perceived the intervention, and (c) the contexts that mediated this relationship. As shown in Table 7.4, qualitative and quantitative data collection methods included purposively sampled semi-structured interviews; direct observation and coding of intervention activities; conversations with the intervention designers, implementers and providers; and participant feedback forms. These are described in detail in the SPIRIT process evaluation protocol.[384]

The research group (which comprised the intervention designers, implementation team, and process evaluation team working in parallel) used the relatively lengthy intervention period as an opportunity to identify, assess and refine hypothesised essential elements during the trial. This was aided by the multi-agency, stepped wedge design of the trial which allowed us to monitor the entire intervention in some agencies and still have scope to trial revisions in other agencies. A modified version of this approach could be applied to other trial designs.

The provision of a dedicated process evaluation researcher as part of the wider group enabled the collection of multiple forms of evaluative data from all sites, and iterative conversations with the intervention designers about their conceptualisation of the intervention’s causal pathways. This allowed us to assess the validity of the essential elements using a five-stage process: Stage 1. Identify provisional essential elements,

---

23 The process evaluation was led by AH but supported by a small group of multidisciplinary researchers based in other organisations. This group were involved in data analysis and reporting.
Stage 2. Test provisional essential elements in intervention contexts, Stage 3. Refine provisional essential elements and develop likely essential elements, Stage 4. Test likely essential elements in intervention contexts, Stage 5. Refine the likely essential elements and develop final essential elements. See Figure 7.3 for a visual overview of this process. Each of these stages is now described.
**Table 7.4. How we answered the three questions for assessing essential elements during the intervention period**

<table>
<thead>
<tr>
<th>Questions used to critique essential elements</th>
<th>Data sources</th>
<th>Data examples</th>
<th>Data analysis / use</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. When implemented in these contexts, does this provisional / likely essential element realise the change principle(s) that informed its development?</td>
<td>Implementation checklist completed during the delivery of each session</td>
<td>Codes showing whether or not (or to what extent) each essential element was delivered as intended</td>
<td>Collation of codes by session and by agency</td>
</tr>
<tr>
<td>2. Is this essential element critical for achieving the session goals? Does anything else appear to be?</td>
<td>Fieldnotes made during observation of each session</td>
<td>Description of how the essential elements appeared to work or not (e.g. how participants reacted), how they were delivered, any adaptations that took place, any factors that appeared to affect how the intervention was delivered or how people engaged with and responded to it</td>
<td>Data was coded thematically using the constant comparative method. In each session we examined the alignment between 1. what was delivered (including any modifications), 2. any observed process effects, and 3. the change principles that informed what was intended, and compared this across all agencies</td>
</tr>
<tr>
<td>3. Does this essential element function across all subcomponents and all six trial intervention settings?</td>
<td>Participant feedback forms collected at the end of each session</td>
<td>How participants assessed delivery against quality criteria such as content relevance, provider credibility, and learning outcomes; and their advice for improvements</td>
<td>Descriptive analysis of quantitative data (frequencies, averages and comparisons)</td>
</tr>
<tr>
<td></td>
<td>Transcripts of semi-structured interviews with purposively sampled participants from two phases of interviewing: early in the intervention period and after it</td>
<td>Participant perceptions of the strategies used to effect change: the extent to which they worked and how modifying factors such as work practices, organisational goals, and beliefs about research shaped process effects</td>
<td>Managed using Framework Analysis. Data was synthesised in categories that were identified both inductively from early interviews and a priori based on intervention outcomes and a review of the research utilisation literature</td>
</tr>
<tr>
<td></td>
<td>Fieldnotes documenting informal conversations with participants following sessions</td>
<td>As above but ad hoc and generally very brief</td>
<td>Data was collated in running memos and, where appropriate, coded thematically using the constant comparative method</td>
</tr>
<tr>
<td></td>
<td>Memos documenting conversations with intervention implementers and providers</td>
<td>Implemeneters’ views on discrepancies between what was intended and what was delivered. Providers’ accounts of why they ‘went off script’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Memos documenting consultations with the intervention designers</td>
<td>How the designers envisaged the change principles manifesting in intervention sessions</td>
<td></td>
</tr>
</tbody>
</table>
Results

These results overlap with our methods in that we show how process evaluation data collection and analysis was used to critique essential elements. This detail is provided so that the procedure we devised is transparent and replicable.

Stage 1. Identifying provisional essential elements

SPIRIT drew on diverse literature and expertise in its design. As shown in Figure 7.3, this body of knowledge was distilled by the intervention designers into an action framework (Figure 7.2) and a list of change principles (Table 7.3) [227, 375, 384] which formed the theoretical basis that we attempted to operationalise in response to each intervention session. These sessions were developed by the intervention designers in consultation with agency staff and expert providers.

Figure 7.3. Process for identifying, testing and refining essential elements (EEs)

We could not use SPIRIT’s change principles as our essential elements. Doing so may have been appropriate for a very flexible intervention with minimally specified, non-standardised components. [55] In such a case, fidelity assessment could focus less on specific operationalisations of the change principles and more on if and how the change principles were realised. [18] However, this was not appropriate for SPIRIT which sought a balance of standardisation and flexibility within a menu of predefined components. The process evaluation aimed to report on variation in the delivery and response to each of these components, consequently the change principles were too abstract to be used as indicators for fidelity reporting. Similarly, the action framework, which functioned as our logic model, outlined causal pathways and relationships in relation to individual and organisational capacity building, but did not identify techniques. We needed a concrete and observable expression of what was at the heart of these strategies if we were to identify commonalities and differences in implementation that could help interpret the outcomes and inform further interventions.

The approach we devised was to identify potential essential elements inductively. As each session outline became available the process evaluation team asked three questions: (a) What do the session goals and the

Questions for assessing essential elements:
1. Do they realise the change principles?
2. Are they critical for achieving the session goals? (and what else appears to be?)
3. Do they function across all subcomponents and all intervention settings?
planned characteristics of the session tell us about which change principles this session is attempting to utilise?, (b) Which of these are likely to be essential to the effectiveness of the session?, and (c) What would these change principles look like in delivery (how can we operationalise them so that can be measured or fully described)? This produced a list of draft essential elements that we further developed with the SPIRIT designers to accurately describe the elements they believed were essential for that session to be effective. These potential essential elements included session content, key messages, provider characteristics, presentation techniques, activities, and particular attendees and types of participation. At this stage, we consciously trialled many essential elements that we suspected would be collapsed or discarded later. See additional file in Appendix 11. for an example.

Devising potential essential elements also required the operationalisation of some relatively abstract overarching concepts. We describe the development of one of these—the concept of quality— in more detail. This is because it is particularly important for ensuring that intervention objectives are achieved [444], yet is neglected in the literature [446, 484].

As per Dusenbury et al.’s definition of quality24 as “the extent to which a provider approaches a theoretical ideal in terms of delivering program content” [444: 244], we conceptualised quality as congruence between (a) the intervention as implemented and (b) the intervention theory—in particular, the change principles. The change principles were strongly informed by adult learning theory which provided quality constructs such as: the providers’ content expertise and presentational skills; the extent to which participants found workshops to be interesting, engaging and respectful of their contributions; the relevance and potential usability of the information and ideas provided; and if participants were facilitated to explore how information and ideas might be applied in their work settings. [485, 486]

We were able to operationalise some aspects of these quality constructs and so include them as evaluator-coded essential elements (e.g. by devising criteria for ‘content expertise’ and using observations to determine the extent to which information and ideas were discussed in relation to participants’ work). However, because quality is highly situated [446], we considered many aspects would be best assessed by participants themselves. Therefore, items in the participant feedback forms were used to collect information about quality constructs such as content relevance, provider suitability, how engaging the session was, and the usefulness of information provided. Quality across the whole program was also considered as part of the semi-structured interviews that were conducted with participants after the intervention. Interviews focused on capturing the breadth of quality criteria from participants’ perspectives (we were mindful that our notion of quality might not align with theirs), and exploring reasons for their judgement rather than ratings.

**Stage 2. Testing provisional essential elements in intervention contexts**

During the first step of SPIRIT (in which the intervention was fully implemented in two agencies and partially implemented in a further two), the process evaluation team monitored adherence to the essential elements, but also gathered qualitative and quantitative data that would help us better understand their real world functionality and validity. We conceptualised validity as: 1. How well the essential elements embodied and delivered the intervention’s theoretical foundations [440, 443, 487], and 2. The extent to which the essential elements were actually essential in each setting [450] (we were aware that elements which seemed essential in one context might not be so in all contexts and circumstances. [447]) Data was collected via observational field notes, checklist coding, post-session memos, participant interviews, participant feedback form ratings and comments, and conversations with providers and implementation team members.

---

24 This definition was used because it is one of the few mentions of intervention quality that is defined in the process evaluation literature
We were attempting to identify essential elements from two perspectives: that of an outside observer (etic) and that of an insider within the group receiving the intervention (emic). In the former we were pursuing our understanding, while in the latter we were attempting to learn from and use participants’ assessments.

During the concurrent data collection and analysis process we adopted a stance of ‘naïve curiosity’ in relation to the essential elements, asking: “What seems to be more or less successful in meeting the goals of each session, and why?” This enabled us to stay open to potential essential elements that we may have failed to consider prior to the evaluation. For example, we noted early on that participants appeared to engage more with session content and gave more favourable feedback when the provider explicitly recognised the challenges of their work, including having a realistic view of the (limited) role of research within it. When the reverse was observed (participants disengaging because the provider appeared insensitive to this issue) we concluded this concept was an essential element of the relevant components: ‘Provider demonstrated sensitivity to the ‘real world’ of the agency’s policy/program work’.

To address our concern about validity we also asked: “How well was the theory underpinning the intervention realised in the delivery of this session?” and “Does each putative essential element appear to be critical for achieving the session goals?” Data was synthesised in running memos that identified issues to explore in further sessions. Analysis focused on comparing our data with the program logic and, primarily, with the change principles that had been identified as informing each session plan.

Six (often overlapping) challenges to the validity of the essential elements were identified through this inductive process. Essential elements could be: 1. Redundant - the element was not essential. 2. Poorly articulated - unclear, too specific or not specific enough. 3. Infeasible - it was not possible to implement the essential element as intended. 4. Ineffective - the element did not effectively deliver the change principles. 5. Paradoxical - counteracting the goals of the session or the underpinning change principles. Or 6. Absent or suboptimal – we identified additional or more effective ways of operationalising the change principles. See Table 7.5. for examples.

Detailed notes were made about the nature of the problem, what interactions affected it (where this was appropriate) and possible solutions that took account of our growing appreciation of contextual constraints and opportunities. Notes included suggestions about where session-specific essential elements could be collapsed and rephrased so that they could be applied across all agencies and intervention components.

Stage 3. Refining the provisional essential elements and developing likely essential elements

The process evaluation team used these notes to amend, distil or reject the 50+ provisional essential elements initially used across the intervention into a list of 26 ‘likely’ essential elements. Following consultation with the intervention designers, these were further revised. The list represented a revised way of articulating and evaluating the fidelity of the intervention but did not affect its design or continuing implementation (with the exception of providers who were sent a list of the essential elements and feedback form items prior to their sessions).

In the revision process, we sought to balance the need for more loosely specified essential elements (which the flexible aspects of the intervention design demanded) with the need to clearly describe what the intervention comprised: not only for the purposes of fidelity assessment but also to provide detailed information that would aid transparent reporting of and potential replication of the intervention. We were guided by Century, Rudnick & Freeman’s account of reducing the granularity with which their essential elements were defined and measured. [480] Consequently, essential elements that had been devised for topic specific sessions were articulated at a higher level of abstraction. For example, ‘The provider demonstrated the value of using systematic reviews in policy/program decision-making’ became ‘The value of
Using research/evaluation in agency work was conveyed. This was necessary because agencies were able to choose and tailor different sessions from within the same intervention component. So in order to monitor fidelity comparatively across all agencies, the essential elements needed to be applicable to every session. Where agencies were able to choose the topic, content, form and goals of face-to-face sessions, the fidelity assessment no longer specified any of these attributes, only that they must reflect the relevant change principles for that component (e.g. those specifying interactivity, shared problem solving, and recognition of participants’ expertise).

Stage 4. Testing ‘likely’ essential elements in intervention contexts

In this stage we used the likely essential elements in our fidelity assessment data collection and continued using the methods described in Stage 2 to collate information about the extent to which they were delivered, and to explore their functionality and congruence with the program theory.

Stage 5. Developing final essential elements

Several further changes were made in this stage but, with some exceptions, not as a result of additional information gathered in Stage 4. Rather the iterative process of refinement allowed us to reflect on details that had been sidelined by more pressing concerns in the previous stages. Having addressed those, we had capacity to focus on less critical amendments and fine tune some essential elements that might otherwise have been considered ‘good enough’. Our final list of essential elements was reduced to 20 items (Table 7.6.). These included several that we considered collapsing but decided to retain separately. For example, is this provider-related element: ‘The provider encouraged participants to contribute to session’ really essential when a participation-related element: ‘Participants contributed to session’ addressed the same concept? Based on empirical evidence from the trial, we concluded it was important to differentiate between (and learn from) what was delivered and how people responded. Our observational data showed that in most sessions the providers’ actions appeared to shape the levels and types of participation, but this was not always the case. Also, because providers were given a loosely specified briefing regarding delivery techniques, as befitted the senior experts who were recruited, we felt it helpful to retain the item for instructional purposes.

Scoring the essential elements

Not all fidelity criteria can be assessed in the same manner. Structural items such as participant attendance, and the number, type and duration of sessions are easily observed and can usually be captured numerically. However, process items (which may be more significant in terms of intervention effects) such as presentation styles, types of participation, and overall quality, tend to be more descriptive and usually require context-sensitive qualitative assessment, especially direct observation. Most of our essential elements were processual so we found that their inclusion in the fidelity assessment required that they be monitored not only in terms of whether they were delivered, but the extent to which they were delivered and how this was done. Our aim was to devise a pragmatic method of standardising observations across sites that could accommodate local adaptation and extensive data collection.

We made three primary adjustments to the scoring as a result of the testing. First, we rejected dichotomised scoring on many items in favour of an ordinal scale. Not surprisingly, we found the Yes / No format we trialled too reductive for the complex processes we were observing. We also trialled several five-point scales (as recommended by Bond et al.) but settled on a four-point descriptive scale of Extensive | Moderate | Limited | Not at all as providing the necessary breadth and precision for our purposes. The definitions that specified the conditions under which each score was applicable were refined in consultation with the
intervention designers and the scale was tested in each agency by two members of the team. All coding was supplemented with descriptive notes.

Second, we developed a scale that could be applied to each customised session (workshop, symposium, etc) and would thereby enable us to compare session content scores across the whole trial. Content was considered to be the aspects of the session that the participating agency had specifically requested. Depending on the nature of the session and the level of detail each agency chose to specify, this content varied tremendously from concrete deliverables (e.g. An example of a systematic review was provided) to relatively abstract processes and concepts (e.g. Ethical challenges were explored interactively). The number of content items also varied from between three to eight. We kept the Yes / No score for each individual item and simply aggregated these using a scale of Wholly | Mostly | About half | Limited | Not at all for each session. This allowed us to compare the delivery of varied content across all sessions and sites without the requirement for a consistent number of items.

Third, we concluded that we had been unsuccessful in finding semi-objective generalisable ways of scoring certain quality concepts (e.g. Was the presentation engaging? Was the content relevant?). We decided to rely entirely on participant feedback to score these essential elements. See Table 7.6. for an overview for the final scoring.

We had sufficient data (checklists, descriptive notes, memos and audio recordings) from the intervention implementation in Stage one to apply these new codes retrospectively to the sessions that informed them.
### Table 7.5. Challenges to the validity of essential elements for the SPIRIT process evaluation and suggested responses

<table>
<thead>
<tr>
<th>Challenges: the putative essential element was...</th>
<th>Definition</th>
<th>Essential element example</th>
<th>Comments</th>
<th>Suggested response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redundant</td>
<td>The strategy described by the element was not essential</td>
<td>'The provider encouraged participants to ask questions'</td>
<td>This was unnecessary in discussion-based sessions where participants interacted as co-contributors</td>
<td>Remove this element</td>
</tr>
<tr>
<td>Poorly articulated</td>
<td>The element description was unclear, too specific or not specific enough</td>
<td>'The session was introduced by a leader (senior person in the agency e.g. CEO, member of executive)'</td>
<td>This failed to capture the many times that less senior staff introduced sessions that were attended by leaders. This essential element was a proxy for visible endorsement/support (modelling) by organisational leaders which we concluded was also achieved when they attended and contributed enthusiastically to the session in other ways</td>
<td>Hone the description so that it accurately captures the essential element</td>
</tr>
<tr>
<td>Infeasible</td>
<td>The essential element described a strategy that was not possible to implement as intended</td>
<td>'Participants were facilitated to identify one or more change goals'</td>
<td>We found this was achievable only in agencies that had developed a research utilisation reform agenda prior to SPIRIT and felt able to use intervention sessions to discuss their goals openly. Other agencies needed more time and different processes to identify goals</td>
<td>Modify or develop alternative strategies. In some cases, the outcomes themselves may need be modified</td>
</tr>
<tr>
<td>Ineffective</td>
<td>In practice, the strategy described by the essential element did not effectively deliver the change principles</td>
<td>'The provider had experience presenting to policy/program developers'</td>
<td>This seemed intuitively reasonable as one of several criteria for securing providers with the expertise and credibility stipulated by our change principles, yet there was no correlation between this criterion and our evaluation of session quality or general participant satisfaction feedback</td>
<td>Consider whether this element can simply be removed or if the change principles require further operationalisation to capture an essential aspect of the intervention</td>
</tr>
<tr>
<td>Paradoxical</td>
<td>When implemented, the strategy described by the essential element counteracted the session goals or the change principles</td>
<td>No examples of this were identified</td>
<td>Interventions can have counterintuitive impacts. While the process evaluation identified examples of this in other aspects of the trial, none related specifically to the essential elements</td>
<td>Remove this element and consider possible implications for other parts of the intervention</td>
</tr>
<tr>
<td>Absent or suboptimal</td>
<td>Additional or more effective ways of operationalising the change principles were identified</td>
<td>'The provider persuasively articulated his/her commitment to using research'.</td>
<td>Despite being briefed to do so, many providers did not articulate their commitment to using research. However, some used case examples that powerfully illustrated the value of research, and facilitated discussion that enabled participants to express it themselves. This strategy was more sophisticated and a better fit with the adult learning orientated change principles that emphasise interactivity, shared reflection and harnessing participant expertise</td>
<td>Introduce absent elements and modify sub-optimally operationalised elements so that the essential aspects of the intervention are captured</td>
</tr>
</tbody>
</table>
Table 7.6. Overview of SPIRIT’s final essential elements: their scoring, how they were monitored, and which of the interventions components they applied to

<table>
<thead>
<tr>
<th>Final essential elements *</th>
<th>Final scoring of essential element</th>
<th>Activity that provided data for scoring</th>
<th>Intervention components to which essential elements apply</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Audit &amp; feedback</td>
</tr>
</tbody>
</table>

**Provider: the characteristics of the person(s) providing the session and their briefing**

1. Provider had expertise and credentials in the topic/field appropriate to the session  
   - Yes / No  
   - Review of publicly available biographical information and, for no.1, participant feedback form item  
   - ✓ ✓ ✓ ✓ ✓

2. Provider had experience in presenting to policy/program developers  
   - Yes / No  
   - ✓ ✓ ✓ ✓ ✓

**Engagement and facilitation: the methods used to deliver the presentation and encourage participation**

3. Non-didactic presentation strategies were used  
   - Extensive | Moderate | Limited | Not at all  
   - Direct observation of session delivery  
   - ✓ ✓ ✓ ✓ ✓

4. Content was delivered in an engaging manner  
   - Yes / No  
   - Participant feedback form item  
   - ✓ ✓ ✓ ✓ ✓

5. The provider encouraged participants to contribute to session (ask questions, make comments, provide examples, participate in discussion)  
   - Extensive | Moderate | Limited | Not at all  
   - Direct observation of session delivery  
   - ✓ ✓ ✓ ✓ ✓

6. The provider encouraged participants to discuss how information / learning from the session might be applied in their setting  
   - Extensive | Moderate | Limited | Not at all  
   - Direct observation of session delivery  
   - ✓ ✓ ✓ ✓ ✓

7. Provider showed respect for participants’ contributions and work  
   - Extensive | Moderate | Limited | Not at all  
   - Direct observation of session delivery  
   - ✓ ✓ ✓ ✓ ✓

8. Provider demonstrated sensitivity to the ‘real world’ of the agency’s policy/program work  
   - Extensive | Moderate | Limited | Not at all  
   - Direct observation and participant feedback form item  
   - ✓ ✓ ✓

**Content: key topics, messages, activities and resources**

9. Core content outlined in session plan was delivered  
   - Aggregated rating across all items specified in session plan: Wholly | Mostly | About half | Limited | Not at all  
   - Direct observation and multiple participant feedback form items  
   - ✓ ✓ ✓ ✓ ✓
10. The session content was relevant to the agency’s work | Yes | No | Participant feedback form item | ✓ | ✓ | ✓ | ✓
11. Where specified in the session plan, provider identified or provided resources that supported or extended learning from the session | Yes / Partially / No / N/A - not specified in plan | Direct observation of session delivery | ✓ | ✓ | ✓ | ✓
12. The value of using research / evaluation in agency work was conveyed | Yes | No | Participant feedback form item | ✓ | ✓ | ✓ | ✓
13. Synthesised data from measures was provided and discussed | Yes | No | Direct observation of session delivery | ✓ |
14. Opportunities to improve use of research were identified | Extensive | Moderate | Limited | Not at all | Direct observation of session delivery | ✓ | ✓ |
15. Targeted agency staff attended | Numbers and roles of all attendees. Approximate proportion of those targeted | Direct observation and review of data from session ‘sign in sheet’ | ✓ | ✓ | ✓ | ✓
16. A leader (e.g. CEO, member of executive) introduced the session or contributed to it positively in other ways | Yes | No | Direct observation of session delivery | ✓ | ✓ |
17. Participants contributed to session (asked questions, made comments, participated in discussion) | All | ~ ¾ | ~ ½ | ~ ¼ | Few | None | Direct observation of session delivery | ✓ | ✓ | ✓ | ✓
18. Participant contributions included knowledge/examples from their own experience | Extensive | Moderate | Limited | Not at all | Direct observation of session delivery | ✓ | ✓ | ✓ | ✓
19. Discussion included how information/learning from the session might be applied in their setting | Extensive | Moderate | Limited | Not at all | Direct observation of session delivery | ✓ | ✓ | ✓ | ✓
20. Participants identified one or more agency research-related areas that could benefit from improvement | Yes | No | Direct observation of session delivery | ✓ | ✓ |

* Essential elements are one type of fidelity criteria. Other fidelity measures concerning frequency, duration, coverage, etc., plus participants’ perspectives, were collected for each session but are not shown on this table.
‘Prohibited’ elements

During the trial we eschewed the concept of “prohibited” [443] or “forbidden” elements [488], but when reviewing the data for Stage 5 revisions we concluded that they could have provided clearer guidance for our providers about the intervention’s underpinning principles. These providers were experts in their field but newcomers to SPIRIT. Despite receiving the essential elements for their sessions in advance, many appeared to apply them selectively. Based on participant feedback and our observations, the following guidance may have helped providers avoid the most common pitfalls:

To be avoided:

- Talking down to participants. In particular, failure to recognise their expertise and the complexity of their work.
- Talking at participants. Didactic presentations should be interspersed with case examples, activities, discussion, etc. Invite questions, ask participants about their views and experiences, and encourage debate.
- Reliance on data/cases from other fields. When information is highly relevant it is more applicable. Where possible, use case examples from the agency’s own work. We can provide assistance with this.
- Squeezing out time for discussion. We conceptualise discussion as a primary mechanism for helping participants integrate new knowledge and think about how it might be applied in their contexts.

We did not trial this guidance, partly because it would have radically changed the provider briefing protocol, and partly because of the potential to alienate eminent highly skilled professionals with such censorious (and potentially patronising) guidance. However, we believe that our methods for assessing essential elements, combined with sensitive consultation with the providers, would glean valuable information about the appropriateness and utility of such an approach. Although this paper concentrates on critiquing and revising essential elements in situ as a means of improving validity in novel contextualised trials, where threats to validity can be identified in advance they should be addressed before the intervention is underway.

Discussion

Identifying an intervention’s essential elements and monitoring them via fidelity assessment is critical for understanding how the intervention worked or why it did not work. Yet there is uncertainty about how to do this, particularly for novel contextualised interventions (i.e. interventions that blend theories pragmatically and which are designed to be flexible and at least partially responsive to local conditions). [442-444, 454, 480] How do we determine which elements of such interventions are genuinely essential to their effectiveness? And how do we ensure they are valid indicators of the intervention theory? [440, 446, 448]

When attempting to answer these questions we found little practical guidance in the literature, and encountered paradigmatic differences and ambiguous terminology. For example, what we call essential elements [444, 481] are also known as essential functions [18], essential components [446], essential ingredients [483], active ingredients [440, 441, 445], critical ingredients [455], critical components [480] and core components. [457, 468] More importantly, they are not always referring to the same phenomenon and they differ greatly in terms of their relationship to the intervention’s theoretical underpinnings. Some refer a-theoretically to intervention activities [446], others to theoretical functions [18]; some use the term to include the breadth of fidelity criteria (e.g. intensity and reach) [454], while others limit it to carefully mapped and validated indicators of theory-based models [489] or recommendations. [450]
Meanwhile, the perceived value of assessing standardised interventions using universal fidelity criteria is declining. The growth of contextualised interventions mirrors increasing recognition of the complexity of the dynamic real world systems in which they are implemented, and the idiosyncratic and unintended ways that interventions and their context can change one another. [18, 23, 25, 68] The need to figure out what fidelity means in such interventions, and to devise methods for identifying and monitoring elements that are genuinely essential, is more pressing than ever.

In this paper we describe a novel exploratory incremental test-and-refine process devised to strengthen the validity of a contextualised intervention’s essential elements. This pragmatic approach enabled us to collect fidelity data throughout the trial (despite uncertainty about what the intervention would look like when implemented in each setting), while also assessing how well the intervention’s real world delivery aligned with the theoretical principles that underpinned its design. The literature provides advice for articulating factual, precise and targeted fidelity criteria prior to the intervention [e.g. 455] but to ensure our essential elements were valid we needed to attend to the interplay of the intervention theory and design with the intervention settings, providers and participants. This was best done empirically in the context of the trial.

Although we monitored implementation fidelity, our methods focused on understanding the intervention’s theoretical fidelity because, as Hawe argues, “Fidelity resides in the theory of the change process, rather than in any particular technology, component, or delivery channel per se. Thus, the role and meaning behind a particular component, rather than its face value, are what matter.” [69:313]

Identifying the appropriate level of specificity was a critical aspect of determining the essential elements’ validity. Overarchingly, we moved from a tightly specified approach to one that was more loosely defined, better reflecting the intervention’s scope for expert providers to shape activities, and for tailoring to individual sites. We knew that session-specific essential elements would need to be distilled into higher order items that covered whole components of the intervention, but testing the functionality and theoretical congruence of a wide variety of provisional essential elements in multiple sessions and sites enabled us to: explore a breadth of possibilities about what mattered and why, increase our understanding of which intervention elements genuinely appeared to be essential, and experiment with how best to articulate and score them. One outcome of this was to increase the extent to which participant feedback was used to measure quality indicators. This approach accords with calls in fidelity assessment, and in research and evaluation more broadly, to use loosely specified evaluation methods that support local adaptation and which recognise that change processes in complex systems are unpredictable and are often best assessed by those receiving the intervention. [18, 441, 470, 482] While none of the process evaluation data, including the evolving fidelity assessment described in this paper, was fed back into the design or implementation of the intervention during this trial25, our approach has potential to contribute formatively to developmental evaluations that shape the intervention during its delivery. [172]

Our fidelity data will be analysed in relation to participants’ feedback form ratings for each intervention session. We anticipate that sessions with higher implementation fidelity will receive a higher overall score and more favourable free text responses. It will not be possible to disentangle the implications of fidelity results for individual sessions or components when analysing intervention outcomes as they are thought to function interdependently, but our data will tell us the extent to which the operational and theoretical aspects of the SPIRIT intervention were delivered in each agency. This, in turn, will help us interpret the observed effects of the overall intervention-as-delivered on outcomes.

---

25 This was because the SPIRIT investigators were trialling a program that was standardised across time and intervention sites.
The use of mixed data collection methods and sources (triangulation), including direct observation and participant feedback, strengthened the rigour of this work. [443, 452, 455, 483] However, the final recursive loop (Stages 4 and 5 as described in the methods section) could have been avoided if we had scrutinised all the essential elements with equal emphasis in earlier steps rather than focusing on those with evident problems.

We note that this approach would not be appropriate for all interventions. Given that the modifications mostly either collapsed essential elements or articulated them at a less granular level, we were able to use the data gathered during earlier phases of implementation to apply the modified elements and codes to the sessions that informed them. However, where essential elements are revised to become more granular (as might be the case in standardised programs where highly specified techniques are being honed), our records would not have contained sufficient detail with which to apply codes retrospectively.

There are other limitations. Our lack of independence as members of the wider study team may have affected our ability to observe the intervention implementation dispassionately and, as is always the case, our theoretical and disciplinary allegiances may have skewed what we noticed and how we assessed it. Lastly, what we observed was situational: shaped by the complex interaction between the intervention theory and structure, delivery by multiple providers, diverse participants and distinct organisational contexts, all at a particular time points. So, while we believe we have identified elements that are at the heart of the intervention theory, we cannot claim that they will necessarily have equal functionality and validity in all settings and circumstances, particularly where they are expressed with greater specificity. [18, 483] We have, however, honed a list of essential elements that appear to be valid in the context of this trial, and which may provide a starting point for others for interventions similar to SPIRIT.

Conclusion

This paper describes the difficulties in identifying the essential elements of a contextualised intervention (i.e. an intervention that is informed by composite social and psychological theories and which incorporates standardised and flexible components in order to maximise effectiveness in complex settings). A worked example of an approach for critiquing the validity of essential elements is provided, including a demonstration of how they can be refined during a trial without compromising the fidelity assessment. This process takes intervention evaluators closer to making theoretically and contextually sensitive decisions upon which to base fidelity assessments in trials of contextualised interventions.
Chapter Eight: The role of internal facilitators

“Change agents must advocate and champion the cause”

Dibella, 2007 [490:235]

Overview

This chapter tackles aspects of how and why SPIRIT was received as it was. The following paper describes how the liaison people—internal facilitators who were nominated by their CEOs to coordinate SPIRIT locally—perceived the study and shaped it in their settings, and how this impacted other participants.

The literature on internal facilitators, champions and opinion leaders has firmly established the importance of these roles in implementing and framing interventions, so the characteristics, views and actions of liaison people were of a priori interest in this research. A straightforward thematic analysis was used because most of the data was strikingly and self-evidently explanatory. Very little interpretation or ‘digging’ to unpack underlying causes was required. This is largely due to the apparently uncensored information provided by the liaison people about their beliefs, perspectives and behaviours. They were highly articulate and analytical policymakers who took a professional interest in intervention design, implementation and evaluation, and were keen to advance our understanding of ‘real world’ issues, even when they were critical of SPIRIT. Their feedback was corroborated and further nuanced by other participants, by discussions with the SPIRIT knowledge broker and study director who had contact with liaison people throughout the trial, and by my fieldwork observations.

This paper makes a solid contribution to knowledge about how trials such as SPIRIT can be facilitated, and how researchers can better engage with these key agents. Explanatory information presented in the analysis is used to develop propositions that may help others to identify and support internal facilitators in similar interventions and, potentially, to inform an evaluation framework.
Abstract
This paper explores the enormous variation in views, championing behaviours and impacts of liaison people: staff nominated to facilitate, tailor and promote SPIRIT (a research utilisation intervention trial in six Australian health policy agencies). Liaison people made cost/benefit analyses: they weighed the value of participation against its risks and demands in the context of organisational goals, knowledge utilisation norms, epistemology and leadership support. There was a degree of self-fulfilment (organisations got what they put in), but SPIRIT could not always be tailored to address local knowledge needs. We present nine propositions for identifying and supporting liaison people in similar interventions.

Introduction
Externally designed and implemented organisational change interventions are thought to have a greater chance of success when they are supported by one or more internal staff members acting as facilitators. Such facilitators often manage the administrative tasks associated with an intervention and may be involved in recruitment, consent processes and/or data collection. More importantly, they are social mediators of the ideas and processes central to the intervention. This may involve formal activities such as presentations at staff meetings, but is likely to include ad hoc negotiation and interpretive communication with diverse colleagues and with those implementing the intervention. Thus facilitators are expected to function as persuasive advocates and mediators, using their interpersonal skills and institutional knowledge to deliver and, where necessary, reframe, interventions to maximise their success.

In this paper we build on existing knowledge by describing the attributes, perceptions, contexts and associated behaviours of the facilitators—known as liaison people — of a novel complex trial that was designed to increase the use of research in health policy agencies. We demonstrate that the liaison people (LPs) functioned as critical mediators with profound impacts on how the intervention was shaped and received in each site. We develop propositions from our analysis that provide concrete guidance about how to identify and support liaison people (or related functions) in similar interventions. But first, we provide an overview of the key roles and characteristics of intervention facilitators in general, and then describe the intervention trial that our LPs were facilitating.

Characteristics of internal intervention facilitators
Intervention facilitators are conceptualised in many ways, but the literature draws attention to three predominant types: champions, opinion leaders and boundary spanners. The terms are not mutually exclusive and are often used interchangeably and/or ambiguously [12, 491], but they denote specific
attributes and functions with implications for how change agents are identified, supported and utilised. [12, 492]

**Champions** are internal employees who advocate for organisational change initiatives. Their function is to capture attention and counter indifference by connecting the intervention with organisational goals and values. Champions articulate their vision of the intervention and demonstrate personal commitment to it. [361, 493] This involves risk as the characteristics of the intervention, including its failure or success, will be associated with their judgement and prestige. [492] The literature describes champions variously as people who emerge spontaneously during a new initiative [361, 493, 494], or respond to a ‘champion call’, or are purposefully recruited. [495, 496] Given the need for champions to be genuinely enthusiastic and to be perceived by colleagues as ‘authentic’, some argue that champions should not be formally appointed. [493] Championing tactics vary [12] and are powerfully mediated by interpersonal and contextual factors. [360] This makes it hard to build champions into standardised implementation planning.

**Opinion leaders** are “able to influence informally other individuals’ attitudes or overt behaviour in a desired way with relative frequency”. [324] Although opinion leaders may mobilise members of an organisation through their expert authority or status [13], they can also be “near-peers”: competent and knowledgeable colleagues who have influence partly because they are seen to share the same frames of reference. [324, 360] Opinion leadership is targeted and topic-specific so different opinion leaders may be required for different types and stages of a change process. Thus someone who is strongly influential in one setting under particular circumstances may find their views dismissed in other settings, or under different conditions. [497]

**Boundary spanners** link people, sectors, interests and perceptions. [498] Their strong external relationships expose them to ideas in the broader environment so they may be more open than other staff to new ways of doing things. [12] They can support interventions by building coalitions and bridging gaps in understanding between the organisation and those implementing the program. [498] Such gaps are often exacerbated by lack of disciplinary or industry knowledge: a common concern in researcher/policymaker relationships [78]. Unlike opinion leaders and champions, the role of the boundary spanner is often formalised.

Effective knowledge brokers (those who facilitate exchange between producers and users of knowledge) [499] possess attributes of champions, opinion leaders and boundary spanners. Knowledge brokers support research-informed policy and practice through knowledge management, linkage and exchange, and/or capacity development, requiring credibility, influence, and the technical and communicative expertise necessary to advance knowledge initiatives within and across complex organisational systems. [302, 340, 500] Many of these characteristics also resonate with Kingdon’s concept of policy entrepreneurs [93]: well-connected advocates who drive change at a macro level (rather than at the meso organisational level). They leverage policy opportunities by linking different facets of the political system (aspects of boundary spanning); and combine technical expertise, influential rhetorical skills and political savvy with tenacity and a willingness to devote substantial time and energy to the enterprise (aspects of opinion leadership and championing). [93]

Common to all these functions is the centrality of complex social processes. [340, 492, 501] Key individuals can influence organisational change, but ultimately it is negotiated through consultation and comparison with peers [77, 82]. Thus the attributes of successful facilitators can only be understood in relation to their context. Greenhalgh and colleagues found that champions were a key determinant of organisational innovation, but that “no amount of empirical research will provide a simple recipe for how champions should behave that is independent of the nature of the innovation, the organizational setting, the sociopolitical context, and so on.” [12:615] Attempts to establish criteria for opinion leaders are similarly confounded:
“What makes someone a credible and influential authority is derived not just from their own personality and skills and the dynamic of their relationship with other individuals, but also from other context-specific factors”. [360:745] Those developing the concept of facilitation concur, arguing that facilitators require a tool kit of skills and attributes that can be wielded for different purposes and contexts, but that their most critical expertise may be in fully grasping the requirements of specific circumstances and responding flexibly. Thus high quality facilitation is that which is most appropriate to the needs of a particular change situation. [328, 502]

**The SPIRIT study and the liaison people who supported it**

As mentioned, this paper focuses on ‘liaison people’, internal organisational staff who were nominated to assist with the implementation of Supporting Policy In health with Research: an Intervention Trial (SPIRIT). Six health policy agencies in Sydney, Australia participated in SPIRIT over a 30-month duration. Further details are provided elsewhere. [227, 375, 377, 503]

SPIRIT’s year-long intervention was designed to increase the use of research by staff in health policy agencies. Its components included locally-tailored educational workshops; structured dialogues with experts in research, policy and knowledge brokering; leadership forums focusing on organisational change; the provision of targeted research products and resources; and access to an online information portal. SPIRIT used a stepped wedge cluster randomised trial design in which the intervention was implemented sequentially, with agencies randomly allocated to the time period in which they received the intervention. Outcomes were measured at baseline (prior to any of the six sites receiving the intervention), then at six-monthly intervals, using structured interviews and a self-reported online survey. An in-depth mixed method process evaluation monitored fidelity and explored the interaction between the intervention, participants and context.

The SPIRIT investigators initially considered employing a member of staff within each agency part-time to act as a liaison person (LP). This would have recompensed the LP for their contribution to the study and potentially increased accountability and effort. However, policy colleagues advised that it would be hard to identify staff who would be suitable for (and willing to take on) this dual role, and that shared management would be problematic. Consequently, a more agency-driven approach was used to maximise local ownership of the intervention: the CEO or equivalent in each agency was asked to appoint a suitable member of staff who would act as the LP. This appointment was a requirement of participation in SPIRIT but, due to the diversity of these agencies, there was no stipulation about what attributes the LP should have other than the ability to assist with a range of administrative, decision-making and promotional activities related to the trial.

LPs were provided with a ‘Liaison Person Manual’ that detailed their responsibilities and timeframes (see Table 8.1), and attended a briefing teleconference with the lead investigator prior to the trial. It was hoped that LPs would assist in maximising awareness and enthusiasm about SPIRIT, as well as ensure it ran smoothly.
<table>
<thead>
<tr>
<th>Phase of study</th>
<th>Task focus</th>
<th>Details</th>
<th>Timing/frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-intervention phase</strong></td>
<td>Liaison person briefing</td>
<td>Attend a teleconference with other LPs in which SPIRIT and the LP role are detailed</td>
<td>A one-off teleconference prior to the intervention for LPs already in place</td>
</tr>
<tr>
<td></td>
<td>SPIRIT introductory session</td>
<td>Attend and participate in the introductory session in which the study (and LP role) is explained to staff</td>
<td>A one-off hour-long session preceding the intervention</td>
</tr>
<tr>
<td>Intervention phase</td>
<td>Selection of intervention components</td>
<td>Consult with colleagues/leaders as required to identify optimal components for agency needs</td>
<td>On-going: Starting after the agency receives its audit feedback and finishing when all components are selected</td>
</tr>
<tr>
<td></td>
<td>Identification of agency interests and priorities</td>
<td>Consult with colleagues/leaders as required to identify topics, content and providers that will best address agency needs</td>
<td>On-going: Starting after the agency receives its audit feedback and finishing when all options are agreed</td>
</tr>
<tr>
<td></td>
<td>Intervention activities</td>
<td>Schedule and book resources for intervention activities</td>
<td>Periodically as required over the 12-month intervention period</td>
</tr>
<tr>
<td></td>
<td>Invitations and reminders</td>
<td>Invite colleagues to participate in intervention activities</td>
<td>Periodically as required over the 12-month intervention period</td>
</tr>
<tr>
<td>Data collection (over 36 months starting at commencement of SPIRIT)</td>
<td>Identification of documents and participants for outcome measures (measures are collected every 6 months over 30 months)</td>
<td>Develop an initial list of invitees and contact details for the online survey based on eligibility criteria</td>
<td>The list is developed before measurement point 1, then updated before each of measurement points 2-6</td>
</tr>
<tr>
<td></td>
<td>Invitations and reminders</td>
<td>Send emails to all eligible staff inviting them to participate in the online survey, followed by two reminder emails</td>
<td>Every six months for six measurement periods</td>
</tr>
<tr>
<td></td>
<td>Process evaluation interviews</td>
<td>Participate in one interview early in the intervention and one following it</td>
<td>Interviews last up to one hour and take place approximately 11 months apart</td>
</tr>
<tr>
<td>Other</td>
<td>Other liaison tasks as required (not specified in manual)</td>
<td>Advocate for SPIRIT, be a resource for colleagues, provide advice to SPIRIT team, act as a communication link between the agency staff and the SPIRIT team</td>
<td>Ongoing over the duration of the study (36 months), but likely to be more intense during the 12-month intervention period</td>
</tr>
</tbody>
</table>
Research questions

While we agree that there is no simple recipe for successfully championing an intervention, we argue that understanding critical aspects of the interplay between personal attributes, views, behaviours, context and the nature of the intervention is possible and can help in the selection and support of effective facilitators in similar interventions. Hence, we attempt to answer four inter-related questions:

1. What were the professional characteristics of the people who acted as SPIRIT’s liaison person and how did these affect engagement with and perceptions of the study (process effects)?

2. How did liaison people perceive and promote SPIRIT and with what process effects?

3. To what extent did the liaison people operate as champions, opinion leaders and/or boundary spanners?

4. How can we explain the variation in how liaison people perceived and promoted SPIRIT? Including (a) What was the role of organisational leaders? and (b) What was the role of organisational context?

Methods

In this paper we report data from the SPIRIT process evaluation. Primary data collection methods were:

- semi-structured interviews with the liaison people and purposively sampled staff in each of the six agencies;
- observations of intervention activities (most of which were attended by LPs); and conversations with study staff who were interacting with LPs during the trial (Table 8.2). Analytic memos written after each data collection event were an additional data source.

LPs were interviewed twice: early in the intervention and post intervention. Early interview questions addressed: the LP’s work role and tenure, their views about agency research use, how they came to be the LP, initial impressions of SPIRIT, and predictions for how the intervention would be received in their agency. Post intervention interviews focused on: their experience of acting as the LP, challenges and benefits, how they tackled the LP tasks, factors that affected engagement, any non-SPIRIT activities affecting organisational research use, support internally and by the SPIRIT team, their views of SPIRIT and any feedback about colleagues’ views, any impacts, and improvement advice. In the post-intervention interviews other staff were asked, “The people who took on the role of facilitating SPIRIT in each organisation were quite diverse. In your organisation X acted in that role. How do you think her/his position here or the way she/he approached the tasks involved in facilitating SPIRIT might have affected how people engaged with it?” Prompts were used to explore participants’ views about the attributes, behaviours and impacts of their LP in more depth.

Interview data were managed in NVivo®[504] using Framework Analysis.[505] This allowed us to summarise and categorise the critical dimensions of the data while maintaining links to the verbatim transcripts. Categories were derived from (1) a priori considerations such as the role of organisational leadership and LPs’ characteristics, and (2) constructs developed inductively from the data such as LPs’ perceptions of intervention flexibility and how they integrated LP tasks into their daily work. A later round of analysis was guided by further concepts from the literature, coding for instances of LPs acting as champions, opinion leaders and/or boundary spanners. Observational and memo data was synthesised into schematic case studies which were structured to allow cross-organisational comparison of key dimensions. During analysis, the LP-related interview data was reviewed iteratively against the case studies to contextualise perceptions, relationships and experiences. Data collection and analysis was concurrent, founded on the method of constant comparison where data is iteratively sought and scrutinised in order to develop, nuance and counter emerging hypotheses and explanations.[386] Synthesised LP data and emerging interpretations were reviewed by a small team of multidisciplinary investigators who contributed regularly to the process.
evaluation work. Later analysis was reviewed by members of the SPIRIT implementation team in order to identify any inaccuracies, and so we could consider their views.

Draft findings were sent to the six primary LPs, i.e. the people who acted as LP for the majority of the intervention in their agency. They were asked to comment on the reasonableness of the findings and to inform us of any other views they wished us to consider. We explained that their opinions would be considered and included in the resulting paper, but would not necessarily alter our interpretations. Our purpose was to: (a) provide the primary LPs with an opportunity to contribute to the depiction of LPs in their agency, (b) re-examine our interpretations in the light of potentially challenging insider perspectives, and (c) provide additional data with which readers could critically assess our findings [506]. Unlike conventional member checking this was not an attempt to validate our findings—people may have quite different though equally valid views of the same issues. [507] All six LPs responded. Their views, including an overview of how their comments changed other aspects of this paper, are presented later. Participants’ perspectives on the findings, together with sampling for maximum diversity of stakeholder perspectives, data triangulation and team involvement in analysis, added to the rigour of this work [508], as did our reflexive stance throughout. [509] Ethical approval for the SPIRIT trial and process evaluation was granted by the University of Western Sydney Human Research Ethics Committee, approval number H8970.
<table>
<thead>
<tr>
<th>Data source</th>
<th>Data collection method</th>
<th>Data</th>
<th>Timing</th>
<th>Focus of data collection</th>
<th>Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liaison people</td>
<td>Semi-structured interviews</td>
<td>Digital recordings, transcripts and memos</td>
<td>During the early phase of the intervention</td>
<td>How they became the LP? Actions and experience of the role to date. Initial views of SPIRIT and predictions re staff engagement.</td>
<td>Six current LPs (plus two people who had acted in LP role earlier)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Soon after the intervention ended</td>
<td>Actions and experience of the role. Views about SPIRIT and staff engagement. Advice for improving the intervention and LP support.</td>
<td>All six ‘main’ LPs plus two who had acted in the role</td>
</tr>
<tr>
<td></td>
<td>Informal telephone calls and emails</td>
<td>Notes and email data (summarised in memos)</td>
<td>Throughout the trial</td>
<td>Views/concerns/further reflections about SPIRIT or contextual factors.</td>
<td>Various, depending on agency</td>
</tr>
<tr>
<td>General participants</td>
<td>Semi-structured interviews (purposively sampled)</td>
<td>Digital recordings, transcripts and memos</td>
<td>During the early phase of the intervention</td>
<td>Organisational culture and context, initial views of SPIRIT, LP role in early implementation.</td>
<td>33 total: between 5 to 7 staff in each agency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Soon after the intervention</td>
<td>Experience of SPIRIT, views about agency engagement, impacts of LP actions and attributes, advice for improvements.</td>
<td>43 total: between 6 to 9 staff in each agency</td>
</tr>
<tr>
<td>Intervention sessions</td>
<td>Observations and checklist completion</td>
<td>Digital recordings, fieldnotes, checklist codes and memos</td>
<td>Throughout the intervention phase of the trial</td>
<td>Documenting intervention delivery, describing participation and interactions in each agency, including the role of LPs.</td>
<td>59 observations across all agencies</td>
</tr>
<tr>
<td>SPIRIT staff</td>
<td>Interviews / structured conversations</td>
<td>Fieldnotes and memos</td>
<td>During engagement phase and after mid-intervention feedback</td>
<td>Any information about views and activities of executive staff and LPs that would help to explain interaction with and impacts of SPIRIT.</td>
<td>Three people who engaged with and gave feedback to agencies</td>
</tr>
<tr>
<td></td>
<td>SPIRIT staff meetings and ad hoc conversations</td>
<td>Memos</td>
<td>Throughout trial</td>
<td>Any information about views and activities of LPs and other agency staff that would help to explain interaction with and impacts of SPIRIT.</td>
<td>The five people most involved in delivering SPIRIT</td>
</tr>
<tr>
<td></td>
<td>Collation of emails from LPs copied to SPIRIT staff</td>
<td>Emails (summarised in memos)</td>
<td>Throughout the trial</td>
<td>Verification of LPs’ email communications to staff in their agency about different aspects of the trial (this information was not received consistently so it is not strictly comparable)</td>
<td>Various, depending on agency</td>
</tr>
</tbody>
</table>
Results

While it is impossible to fully disentangle their impact from other contextual factors, it is evident that LPs made a profound difference to the way that SPIRIT was communicated, perceived and engaged with in each of the six intervention sites. We present the findings in relation to our research questions: 1. What were the professional characteristics of the liaison people? 2. How did liaison people perceive and promote SPIRIT? 3. To what extent did liaison people operate as champions, opinion leaders and/or boundary spanners? And 4. How can we explain the variation in how liaison people perceived and promoted SPIRIT? Including (a) What was the role of organisational leaders? and (b) What was the role of organisational context? Some details about LPs and the organisations in which they work have been altered to preserve anonymity. We have deliberately obfuscated any details that may reveal which LP was based in which organisation and, in the case of multiple LPs in the same agency, their position in the sequence and whether or not they were the primary LP.

1. What were the professional characteristics of the liaison people?

In five agencies, the CEOs nominated the most senior member of staff with a research or evaluation role to act as the SPIRIT LP. In two cases, these staff passed on the function to a more junior member of their team. In the sixth agency, which had no dedicated research or evaluation position, the LP function was held by a senior executive. Consequently, there was substantial variation in the hierarchical position and role of LPs within their organisations.

Seniority

Contrary to our expectations, LPs with greater seniority did not always facilitate higher levels of participation. Reminders sent by the most senior LP often resulted in a spike in online survey completions compared to a nil effect from less senior LPs, but there was no consistent relationship between seniority and survey response across the agencies. Interviewees speculated that simply appointing a senior person to act as the LP “spoke volumes” about the “authenticity” of that agency’s commitment to SPIRIT. How this seniority was used, however, was equally important. Leveraging power may have had adverse impacts such as causing resentment from staff who were instructed to attend workshops. Conversely, seniority enabled LPs to make executive decisions, whereas junior LPs needed to elicit responses through bureaucratic chains of command, often negotiating new processes for dealing with the questions SPIRIT posed. This meant they often took longer to complete core tasks, but this did not result in lower participation rates overall.

Organisational role

Interviewees stated that where the LP had a research or evaluation position this bolstered their credibility as an appropriate facilitator for SPIRIT, but in some cases this association constrained how SPIRIT was perceived. For example, in an agency where the LP was an evaluation manager some staff assumed SPIRIT was an evaluation program, “who that person is affects what you think the presentation is going to be about [and its] relevance to your team… because she might have clearly explained [the study] but in your head it’s evaluation”. These interviewees speculated that their LP’s position increased the intervention’s credentials as an evaluation resource, but reduced the likelihood that staff who were not involved in evaluation would participate. In another agency, several participants expressed discomfort about the purpose of the relatively new organisational position that the LP occupied and indicated that this led to negative connotations for SPIRIT.

Liaison person coverage and workload

There was considerable variation in the turnover and coverage of the LP function during SPIRIT. Two agencies had a single LP for the entire study period, but with lengthy absences in one case. Two had a single LP during the intervention phase, but different LPs during data collection. In the remaining agencies,
multiple people acted as the LP across both phases, including periods during the intervention in which there was no LP. This impeded the conduct of the outcome measures and the selection of, and arrangements for, intervention sessions. It also appeared to impact negatively on staff awareness of the intervention with several interviewees in these agencies stating they did not recall any communication from their various LPs about SPIRIT. In most cases, staff turnover or restructuring caused the LP function to be transferred, but in one agency it occurred because two LPs found the demands too onerous. As one manager explained, “It ended up being a bit more work than we’d anticipated... she literally just couldn’t manage it all”.

These demands were not fully knowable in advance, partly because this was a novel trial, but also because LPs’ workloads differed hugely depending on how they consulted about tailoring the intervention. The SPIRIT team tried to minimise the burden on LPs but found that they had underestimated time requirements in some cases.

2. How did liaison people perceive and promote SPIRIT?

Liaison people across the six agencies had strikingly different conceptualisations of the LP function which, in turn, shaped how they approached the tasks. In one agency, SPIRIT was implemented following a major restructure. The LP speculated that her colleagues would conflate her newly formed team with SPIRIT: “… my feeling is that people will tend to judge [us] by how useful they find SPIRIT, but also, maybe, to judge SPIRIT by whether they are embracing [us] or not”. Accepting this blurred line, she focused on adapting the intervention so it could be integrated into the team’s planned activities and directly support their professional development goals. The LP in another agency saw the function as an extension of his research governance position so he used the experience to further develop cross-agency networks, convene research-orientated forums and increase essential skills in “translating and negotiating”. A third LP had operational oversight of the organisation and conceived her core task as managing the study’s demands: “getting it done efficiently”. She used her authority to act as a buffer between the trial and already overburdened staff, and to maximise measurement responses, “if I say ’Do it’, people will do it”. Another LP was appointed to “fix” SPIRIT after a previous LP had failed to engage staff. She conceptualised the work as a mobilisation exercise that depended on “getting buy-in”, so she prioritised interpersonal persuasion and advocacy. The LP in the fifth agency, who managed a research team, argued that the LP function was primarily administrative and questioned how appropriate it was for someone in his role. He focused on the core deliverables and minimised other contact with the trial. Whereas the LP in the sixth agency had an equivalent organisational role but saw the LP function as a “natural fit”. She conceptualised her task as maximising the value of SPIRIT which meant “generating belief” among managers so they would persuade their staff to participate, and devising mechanisms to “embed” the intervention’s ideas in organisational practice, “I tried to get something out of each [workshop] that would stay, would hang around for us.” Thus, in all cases, it seemed that the LP function and its core tasks were perceived in relation to the organisational position and professional responsibilities of the people who were assigned to act as LPs. As we show later, this was further shaped by the study’s perceived alignment with wider organisational goals.

3. To what extent did the liaison people operate as champions, opinion leaders and/or boundary spanners?

Championing SPIRIT

Of the six LPs who were in place for the majority of the intervention phase, four were champions for SPIRIT and two were not. Non-champions did not: communicate a vision of what the intervention could achieve, demonstrate commitment or enthusiasm for it, or attempt to engage others in supporting it. [493] Of the further 13 people who acted as LPs temporarily during the trial, we estimate that four displayed moderate championing, three were clearly not champions, and the other six (some of whom were in position for a matter of weeks) are unknown. We base these assessments on: (a) LPs’ statements in interviews and/or
informal conversations about the value of SPIRIT and how they approached the tasks, (b) observable
differences in LPs' levels of enthusiasm and approach to tasks during intervention activities (e.g. how they
introduced staff workshops), (c) LPs’ conversations with SPIRIT team members, and (d) interviews with other
agency staff. There was close agreement between LPs’ self-reported attitudes and behaviours in relation to
the intervention and how other agency staff perceived their LP.

Unsurprisingly, LPs’ regard for SPIRIT appeared to correlate with their ‘championing’ and this, in turn, had an
impact on how the intervention was structured, promoted, attended and perceived in each agency. In the
two agencies where none of the LPs were champions, there was less consultation, the intervention was
promoted inconsistently, aspects of the tailoring were less successful, and participants’ views of the
intervention expressed in interviews and workshop feedback forms were more negative than in the other
four agencies.

Relationship quality
Congruent with the literature which asserts that the quality of relationships between champions and their
colleagues is positively associated with influence [510], many interviewees reported that positive regard for
their LP encouraged their participation, “[she] is such a wonderful person that you kind of want to do it for
her”. Strategies LPs used to encourage goodwill participation included dropping by colleagues’ desks to
request they complete a survey, chatting in the kitchen about SPIRIT goals, and, in one case, negotiating an
explicit quid pro quo. Where interviewees reported instances of participating as a direct result of their LPs’
request, it was usually in the context of informal personal interaction (rather than emails or generalised
comments in meetings). This suggests that friendly near-peer LPs may be best positioned to encourage
participation. [324] However, as we point out later, leaders also played a vital role in this dynamic.

Selling SPIRIT
One of the most noticeable differences between champion and non-champion LPs was the extent to which
they “marketed” SPIRIT, i.e. creatively harnessed organisational information channels and used rhetorical
strategies to make the intervention and outcome measures more appealing. One LP did a mini presentation
for staff who had missed the introductory session. He admitted to “embellishing the outcomes [feedback] a
little bit” to create buy-in. Another explained, “I can talk it up in a way that sounds like it’s not a hassle and
it’s interesting - and look, it is good stuff we’re getting out of this that will help you in your wor
This LP
sought out and spoke personally to every member of staff nominated for each of the outcome measure
points (about 25-30 people on six occasions), and achieved a 100% response rate.

One of the more senior LPs was also keen to increase response rates so she sent a rare personal memo to
staff telling them that, uncharacteristically, she had completed the survey because the organisation needed
good data for cross-agency comparison. In interview she explained her strategy: if staff knew she had
completed a “bloody survey” they would understand its importance, plus they are highly motivated by
competition. This agency’s response rate increased significantly and we are not aware of other factors that
could explain it. This accords with findings that champions use formal and informal methods of
communication to frame interventions strategically in terms of organisational orientations and objectives.
[361, 493] Conversely, non-champion LPs may have undermined SPIRIT at times by overtly distancing
themselves from the study. For example, one LP introduced a workshop saying that he didn’t know what it
was about, and another forwarded email requests to colleagues about the online survey with the disclaimer,
“Don’t shoot the messenger!” As we show later, these activities were influenced by LPs’ concerns about
SPIRIT impeding their day-to-day work and damaging their professional reputation.

The blurred distinction between persuasion and imposition was noted in every agency. All LPs admitted to
“cracking the whip” to some degree, and most reported that staff sometimes felt hassled by multiple
requests to participate. Overarchingly this related to “trying to get people involved with something that they
don’t necessarily see benefits them directly.” Some LPs argued that getting colleagues to see these benefits placed too much responsibility on them; they asked, “Whose job is it?”. Naturally, LPs who valued SPIRIT were more willing to sell it and to convince colleagues to participate in data collection as part of a trade-off, but they also had more ammunition with which to do so. “Chasing” colleagues required “resilience”, but there is little doubt that these persuasive strategies increased participation; as one of the ‘chased’ participants explained, “people find it hard to say no because... [the LP] is very politely persistent in that she’ll find you and hassle you until you [say yes]”.

These findings highlight the ethical challenges of workplace interventions. As others acknowledge, the line between persuasion and coercion is particularly delicate in organisational research where co-workers have recruitment responsibilities, protocols cannot be easily enforced by the research team, and where staff may regard participation as expected. [511] SPIRIT sought to minimise coercion risks by reiterating the voluntary nature of participating and providing opt out opportunities pre- and post-data collection. Although managers could see who attended sessions they did not know whether their staff participated in data collection. There were no complaints about coercion, and the low survey response rates in most agencies suggest that people did not feel compelled to complete them, but in supporting the study a minority of LPs and managers may have strayed over this line, resulting in some unwilling participants (albeit in a study with negligible risks of harm).

Understanding SPIRIT

Being an LP was a learning curve and many of the longer serving LPs found that they became more adept over the duration of the study. Familiarity with the outcome measures increased the efficiency with which they were managed, just as experience of the workshop consultation and delivery process illustrated how components could be adapted, “[Before this] I couldn’t envisage what a Research Exchange would look like – what the possibilities were”.

Some non-champion LPs were unable to explain the study to their colleagues. One did not know what was happening in the intervention or measurements. Another seemed unaware that his agency had choices about the intervention content: the selection of which he was meant to be facilitating. Many of the LPs found the study information dense and excessive so concluded that grasping it was not a worthwhile use of their limited time. One of them minimised the need to understand the study by telling her staff if they had any questions they should talk to the SPIRIT team. Another handed over the LP function to a colleague when action was required. This contrasts with one of the champion LPs who so delighted in knowing everything about the study that she playfully asked SPIRIT staff to test her on the details.

SPIRIT support for championing

All but one LP described the SPIRIT team as supportive, but several felt the team could have done more to build relationships and anticipate their need for succinct, sharable information. Some pointed out that, ironically, this was a research translation issue. Support was also not always consistent: some LPs who took on the function during the intervention period received less instruction than their predecessors. In one case, the LP felt this impeded her ability to champion SPIRIT,

“You should have really sat me down and said, okay, this is what it’s all about...That would have clarified the whole thing to me and I would have been able to say, okay, I can explain it to everybody and promote it, advocate for it, I suppose, which I don’t feel I have really been able to do.”

LPs made suggestions for improving communications and support, highlighting the need for more on-site visits and face-to-face conversations, particularly in the early stages of the trial.
Opinion leadership

Participant interviewees described the characteristics by which they judged the suitability of their LPs; these spanned championing and opinion leadership and were contingent on two related concepts of legitimacy: credibility and commitment. Colleagues in one agency argued that it “didn’t make sense” for SPIRIT to be promoted by their LP given her seemingly limited understanding of its aims and her indifferent attitude towards research in general. Conversely, colleagues of another LP commented that she was “ideal” given her “research credibility” both as an academic and an enthusiastic advocate for research-informed policy. The LP in a third agency concurred, explaining the CEO nominated him because he is publically “committed to evidence” and known to influence colleagues’ engagement with research. All but one of the primary LPs thought LPs needed research experience in order to speak authoritatively about SPIRIT, and most interviewees regarded knowledge of the organisational culture as critical. There was no indication that LPs either were or were not viewed as opinion leaders in relation to other aspects of organisational business.

Boundary spanning

SPIRIT asked LPs to function as boundary spanners across different parts of their own organisations as well as across the organisation-SPIRIT divide, hence LPs who were newly employed members of staff were at a disadvantage: they “couldn’t leverage existing relationships” or make informed judgements about which colleagues and what documents were eligible for the outcome measures. Lack of familiarity with workplace culture and communication styles complicated consultations about how best to use SPIRIT but, as LPs’ relationships evolved, appeared to have less impact on later phases of the study. For example, a newly employed LP was keenly aware that she lacked essential workplace knowledge; yet, by the time of the post-intervention interviews, she was seen by colleagues as a highly effective networker and “ambassador” for SPIRIT. This LP had used SPIRIT to initiate organisational connections and had formalised boundary spanning by recruiting colleagues across the agency to act as team advocates for SPIRIT.

For an LP to bridge the divide between their organisation and SPIRIT, some “translation” was required. Most of the LPs attempted to make the study terminology and underpinning concepts more accessible, e.g. they interjected during workshops to explain terms and provide illustrative examples. They also provided reassurance such as when, in the more clinically orientated agencies, LPs used a drug trial analogy to illustrate that the intervention was being tested, not the participants.

Mediation

Although the SPIRIT team had mechanisms for communicating to agencies, they were dependent on LPs for conveying communication from agency staff. Consequently, lack of boundary spanning by LPs in some sites meant the SPIRIT team had no access to participants’ views and concerns (the process evaluation did not provide this feedback until after the intervention). Conversely, the more enthusiastic LPs acted as mediators which increased the extent to which concerns were aired, addressed and fed back. For example, when the online survey was shortened one LP framed it as the researchers’ response to criticisms raised by staff in that agency, advising them “if you do have any questions or comments at any time about SPIRIT then you can tell me about them because they are listened to and this is evidence of that”. It seems likely that these staff would have perceived such feedback as a validation of their participation. Staff in organisations with less communicative LPs might have welcomed the shorter survey but would have had less sense of agency in bringing it about.

Brokerage and advice

Variations in boundary spanning resulted in very different levels of advice from the LPs which impacted the SPIRIT team’s sensitivity to each organisation’s culture. Participants in one agency criticised SPIRIT for not using professional learning techniques that were their standard practice (they wanted directed “pre-readings... so we can come into the room with our heads in the correct space”). These could have been incorporated if we had known. A forthcoming LP might have informed us about these norms unbidden, but
we missed an opportunity to learn from the agency prior to the intervention about how to optimise activities in their setting. So while the trial benefitted from boundary spanning LPs who proffered advice and creative suggestions, if we had acted as better boundary spanners ourselves we may have been able to tap into valuable insider knowledge more effectively across all the agencies.

SPIRIT team responsivity
Effective boundary spanning was a two-way street requiring mutual responsiveness and conciliation. LPs identified four behaviours from the SPIRIT team that they found particularly encouraging: 1. SPIRIT staff sending positive reinforcing feedback about the LPs’ hard work and positive impacts to their manager/CEO. 2. Small appreciative gestures from the implementation team such as ‘thank you’ emails, verbal acknowledgements during workshops, and gifts of chocolates at Christmas. 3. Changing aspects of the trial in response to agency feedback (e.g. shortening the online survey), and 4. Supporting LPs to use their expertise to adapt information materials and participation strategies, “the good thing was that [the SPIRIT team] always acted on what I suggested... [they] realised that I know the organisation better than they would and what works here”. Thus the positive interactions between LPs and the SPIRIT team were co-adaptive. Where LPs suggestions were not acted on (usually due to infeasibility or adherence to the study protocol) this caused frustration. Clearer communication about why those decisions were made might have lessened this irritation and provided the LPs with a rationale they could share with colleagues.

4. How can we explain the variation in how liaison people perceived and promoted SPIRIT?

Cost/benefit judgements
LPs made informal cost/benefit analyses about the potential value of the SPIRIT intervention for their organisation in relation to its demands, determining their levels of enthusiasm for the intervention, how they perceived the LP function and how they approached its tasks. For example, one of the champion LPs was explaining her hope that SPIRIT would “pay off”,

“...it’s certainly helped the general direction that we want to travel in terms of the role of research. So in that sense, yes. It’s been fairly time-consuming for me personally, but probably worth it for the organisation”

Perceived costs and benefits were influenced by management attitudes and behaviours, and by other organisational factors as described below, but were also entwined with an assessment of the potential professional benefits and risks in being associated with SPIRIT: those who expressed most enthusiasm about organisational benefits also identified value for themselves in being the LP. This assessment was particularly evident in two agencies in which the LPs were new employees. The one with a positive view of SPIRIT embraced the LP role, anticipating that it would help her develop internal connections and stakeholder relationships that were critical for her day-to-day work. After the intervention she reported that it had done just that. Whereas the other LP tried to minimise the risk that SPIRIT would be perceived as his project, “I didn’t want that connection”. As a new employee with no established organisational reputation it was uncomfortable to be associated with activities that he regarded as demanding with dubious merits. In three agencies LPs saw SPIRIT as a resource that bolstered their extant work in developing organisational research or evaluation capacity and, post intervention, they identified positive impacts in relation to their work. The least enthusiastic LPs did not identify positive impacts for their agencies or themselves. The ‘risk minimising’ LP described above said that the role had “helped expose me and connect me with people”, but not in the manner he would have chosen.
Being nominated

There was no association between how people came to be the LP and their attitude towards it. The only self-delegated LP was among the least enthusiastic. Conversely, the LP who was “volunteered” in her absence went on to engage an overtly disengaged organisation and to facilitate one of the highest proportional attendance and survey response rates overall. Her initial views of SPIRIT as a confusing “research thing” were far from enthusiastic, “[When] I got back from holidays and I was asked to take it over I was, like, ‘Oh my God! Why?’” Despite this inauspicious start, she strove to learn about SPIRIT and became convinced that her organisation could benefit. Motivated by this and the challenge of turning around the previous LPs’ lack of success, she approached the LP tasks with gusto and was able to incorporate a ‘conversion narrative’ as part of her rhetoric, “I’d say, ‘look I thought the same as you… what a hassle!’ But... it’s actually much easier than you think”. This echoes findings that ‘change cynics’ who revise their views of an intervention can become highly effective champions. [495] It is also another ethical grey area in that several LPs were reluctant participants.

a. What was the role of organisational leaders?

Permission to push

Although LPs were asked to be the ‘face’ of SPIRIT in their agency, perceptions of the extent to which they were representing managerial views were critical. Three LPs said they felt justified in being assertive about SPIRIT because it was known to be on behalf of the organisation’s leaders, “They knew it was something that I was pushing, but not for my own agenda... I was nagging them on behalf of our upper management.” In cases where leaders explicitly demonstrated support for the LP’s SPIRIT-related activities, they felt this “imprimatur” was strengthened. Colleagues in these organisations concurred. According to interviewees across all agencies, the most persuasive incentive for completing the outcome measures was being asked by a well-liked, well-respected colleague who saw the endeavour as worthwhile, backed by evident managerial/CEO support.

In contrast, another LP expressed discomfort about the organisational burden of repeated outcome measures and his need to cajole staff to complete them. Despite strong CEO espousal for SPIRIT, managerial support in general was not as visible or consistent as in some other agencies. For example, the LP’s immediate manager expressed scepticism about SPIRIT during workshops. This may have contributed to a less conducive environment for persuading colleagues to participate. Managerial cynicism has been found to depress staff attitudes to organisational change initiatives. [512] Even where LPs perceived managers as committed to SPIRIT, they often struggled to get visible back up, “The main challenge for me, I think, is engaging our leaders enough so that they can convey the message to staff”. And in some cases, managers constrained LPs’ initiatives, limiting SPIRIT’s reach in the process. Examples included refusing an LP’s request to introduce a ‘SPIRIT slot’ at team meetings, and instructing the LP to reduce burden on the agency by limiting the number of staff who were invited to participate.

SPIRIT and work performance

The extent to which SPIRIT was formally recognised as part of the LPs’ work varied. In most cases managerial oversight of SPIRIT was added to the LPs’ usual reporting lines. In three agencies, LPs and their managers identified ways that SPIRIT could be used as an opportunity for professional development, e.g. using LP activities as a vehicle for increasing their status and/or exposure in the organisation, and building SPIRIT deliverables into performance reviews. Such strategies strengthened these LPs’ desire to “make it work”. Where LPs radically shaped SPIRIT to address organisational priorities this was possible only because the LP already had some responsibility for developing such initiatives, and there was managerial support for using SPIRIT this way. Figuring out how to accommodate and use SPIRIT within LP’s work was dependant on managers understanding the scope and responsibilities of the function and how these could be enacted in their organisational context, “It is one thing nominating a liaison person and then another thing to find, oh,
does that liaison person have the authority to take decisions on all of these areas or to speak across the organisation? Or is their role more administrative?"

b. What was the role of organisational context?

Paradigmatic compatibility

Perceptions of SPIRIT’s compatibility with the organisation’s conceptualisation and use of evidence appeared to be the strongest determinant of why LPs saw greater or lesser value in SPIRIT. When interviewed, two of the most unenthusiastic LPs explained that the intervention made assumptions about how they should be engaging with research that did not align with their practice,

"People are operating at a different level from what is assumed [by SPIRIT], and have different needs. It’s no longer to do with access to research evidence, it’s what do you use and how do you use it to articulate good practice? How do you cut through the politics? How do you get people at the frontline to become aware of what they do and get them to throw back at you what kinds of questions are important, and how can that translate into research and policy? Which are very different kinds of questions from just how do you get more research into policy."

These LPs rejected the implication that their organisation should improve their use of research in the way SPIRIT conceptualised it, and did not believe that an externally designed intervention was an appropriate means of tackling highly situated knowledge-to-practice concerns. Their views were supported by other interviewees in the same agencies suggesting that they were representative of their dominant workplace cultures. It is possible that, say, practitioners from other jurisdictions sharing real world experiences, or workshops that focused on internally developed research or evaluation might have been more welcome. But the more disengaged LPs seemed unclear about how much intervention opportunities could be adapted and may not have considered these to be possibilities. In one case, the SPIRIT team pushed for a workshop to be facilitated collaboratively with an expert in that agency. The idea was welcomed in principle, but later dismissed due to work pressures.

The more enthusiastic LPs worked in organisations that saw evidence, or the intervention, in a slightly different manner. Although all agencies had a pluralistic conceptualisation of evidence, an investment in stakeholder engagement, and extensive experience in implementing policies and programs in messy real-world contexts, their emphases varied in accordance with their remit. Agencies working within specific biomedical fields (two of the intervention sites) seemed more disposed to embrace evidence-informed ideals than those with broad population health or systems reform briefs. This may reflect the extent to which forms of research considered to be of highest academic quality—such as randomised controlled trials—could be applied instrumentally in their contexts. However, two of the champion LPs were in agencies with far broader remits. The first of these agencies was directly dependent on Ministerial approval (and therefore, arguably, most susceptible to overt political pressure), yet their LP embraced SPIRIT. Several factors may have played a role. First, there were positive pre-existing relationships between the intervention designers and staff at different organisational levels who had commissioned some of the components offered by the intervention. Having used (and, to some extent, shaped) the product on offer, staff in this organisation were probably less likely to dismiss SPIRIT as pushing a purist and irrelevant evidence-based agenda. Second, the agency leaders enthusiastically and credibly espoused research utilisation and explicitly supported SPIRIT and the LP as a champion of the intervention. These factors were likely to reassure the LP that SPIRIT was sufficiently compatible with his agency to be worthwhile.

The second agency was embarking on training to strengthen their in-house research and evaluation capacity. The LP suggested that their continued participation in the study was contingent on SPIRIT
contributing to this pre-existing agenda, and she negotiated assertively to refashion intervention activities accordingly. Paradoxically, lack of established relationships between the SPIRIT team and agency staff may have facilitated this exchange as the agency had little to jeopardise in taking a strong stance. The commonality in all cases was the need for alignment between SPIRIT and the agency’s current engagement with research.

**Tailoring and alignment**

There was a strong sense of each agency being in flux and striving toward particular practice goals. This trajectory appeared to provide the benchmark against which LPs assessed the value of SPIRIT: *given our circumstances and strategic goals, is this intervention worthwhile? To what extent does it provide opportunities that support how knowledge is conceptualised and situated within our day-to-day practice?* This was echoed by interviewees’ predictions about whether they would participate in SPIRIT. To do so they would need to “see value”, gain “practical benefits”, and know that the intervention had “a direct relationship with the work that I’m doing.”

Program flexibility and responsiveness was a key criterion for this assessment. The least enthusiastic LPs expressed concerns about structural inflexibility, *“The tailoring of the program is not really tailoring. What it is, you’re giving us a menu… you told us what you’re doing… and all we’re doing is ticking the boxes”.* They saw limited scope for extensive adaptations because they regarded SPIRIT as fundamentally non-consultative, *“You’re talking to [us] but it’s a one way situation”.* However, where LPs experienced the intervention as genuinely tailorable they maximised its benefits by working with SPIRIT staff to shape the workshops and resources to address organisational priorities. Two LPs integrated intervention components into a wider program of staff capacity building, selecting topics, content and formats specifically to complement internal initiatives. Timing was also critical. Managers in the organisations in which these two LPs were based wanted the intervention to start at the same point as their internal initiatives, and one insisted on a hiatus while a major restructure was finalised. SPIRIT may have been better integrated by other agencies if organisational leaders could have decided when the intervention commenced.

**LPs’ view about our findings**

All six of the primary LPs who were invited to comment on the manuscript responded. Three gave general neutral or favourable feedback, and three commented more specifically. LPs were asked to alert us to any concerns about their identifiability but none did so (though one was initially concerned that other LPs might be identified). One LP asked for a word to be softened and another questioned an ambiguously phrased description of her agency. We agreed with their feedback and made amendments they were satisfied with. Two LPs developed themes in the manuscript about aspects of the trial that motivated them (belief in the goals of the trial, wanting to work with the trial leaders, leadership support and building LP tasks into their work performance review) and the characteristics required for the LP function (organisational and communicative skills, cross-agency connections and tenacity). This feedback has been included with the findings they relate to. Two felt that, having read what LPs in other agencies were doing, they would have benefitted from interacting periodically during the trial to share experiences and discuss strategies, *“after reading the manuscript, I really felt the loss of not having an opportunity to interact with other LPs – I think we could have learned a lot from each other!”*

**Implications**

Our findings highlight some of the challenges of implementing complex interventions in real world settings where the intervention’s ideas and activities must be carried by, and work through, existing organisational structures, processes and relationships. [13, 360] In such interventions change is a series of entangled interactions which are impossible to fully control. [12] Nevertheless, it is also apparent that LPs wrestled with
practical and, in some cases, conceptual obstacles which, in hindsight, could have been better anticipated by the intervention team. For example, we could not manage the frequency with which the LP function was transferred during the intervention, but we did know that policymakers change jobs rapidly and often act in other roles, so we could have designed a better system for supporting these transitions. Further, LPs were not explicitly selected for their potential to act as champions, opinion leaders or boundary spanners — the need for these attributes was only realised during the intervention rollout when it became apparent that most organisational leaders were too far removed from SPIRIT’s day-to-day activities to act as its social facilitators. Greater attention to the selection of LPs is clearly indicated.

The conceptual obstacles experienced by LPs suggest that more fundamental revisions should be considered. This will be addressed in further papers when all of our trial data can be considered. In the meantime, we acknowledge the dilemma that different views of SPIRIT presented for some of the LPs. From one perspective, LPs’ views were self-fulfilling: where they judged the intervention to have potential value they invested their efforts thereby adding value and experiencing SPIRIT as worthwhile. Where they judged it to have little value, little was added and little was experienced. But how much value could be added? The non-champion LPs were reflecting wider organisational concerns about the dissonance between an externally developed intervention that appeared to pre-frame the problem it was addressing, and in which experts provided generalisable knowledge when they saw knowledge as constituted through local practice: what Gabbay and le May call knowledge-in-practice-in-context. [31]

It may be helpful to consider these findings in the light of previous research that shows even where change agents are highly respected opinion leaders their influence is bounded by current organisational norms and expectations. [324] It is far easier to motivate people who are receptive to the ideas presented in an intervention than those who are cynical [324]. Thus LPs may have been able to galvanise people’s engagement with SPIRIT positively or negatively but, without modifying the intervention substantially, could not have driven transformative change that countered dominant cultural tendencies, no matter how personally committed they were. [360, 490, 495] The existing culture of research use within a policy organisation is known to affect how research utilisation intervention strategies are received. [295]

The findings from this study support those observed in other studies in that the delivery of interventions is profoundly affected by those who act in facilitation roles akin to that of our LPs. [e.g. 328, 513, 514] Further, that the LPs’ ability to function as champions, opinion leaders and boundary spanners, were critical. For example, Dixon-Woods and colleagues found that interventions were most effective when,

“... those locally charged with implementation were sincere in their beliefs about the value of the program, were able to create transdisciplinary alliances, had local credibility among peers, were prepared to tolerate debate but exercise firmness, and used multiple tactics including role modelling, persuasion, sanctioning, reminders, and constant feedback”. [392]

What this study adds is an analysis of how these issues played out in a research utilisation trial in policy agencies. In particular, our findings suggest that concepts from political science about the contingent nature of evidence in policy [33, 90, 119, 515] and how it intersects with policy practices and organisational change [83, 133, 516, 517] were central to how the intervention was facilitated. The paradigmatic compatibility of SPIRIT with agencies’ current and proposed research use strongly affected each LP’s views about the value of SPIRIT and this shaped how they engaged with and facilitated the intervention.

A standardised checklist of LP attributes is not meaningful in isolation, but we believe that some propositions (generalised theoretical statements grounded in the data [505]) can be drawn from our findings. Given the complexities outlined above these propositions may be somewhat aspirational, but they point us in the right direction for identifying and supporting LPs in interventions similar to SPIRIT and,
potentially, for informing a framework for evaluating attributes and conditions. They are clustered in three categories: Liaison person attributes, Managerial support and Intervention team responsibilities.

Liaison person attributes

→ Proposition 1: The LP must believe that the intervention is worthwhile. At best, they will be genuinely enthusiastic about its merits – a champion. At least, they will judge that the benefits outweigh the demands.

As expected, the ideal internal facilitator for an intervention study such as SPIRIT appears to be a genuine champion (someone who believes in the intervention and will advocate for it energetically), an opinion leader (someone with informal organisational influence), and a boundary spanner (someone well-networked in their workplace who can also communicate effectively across the intervention-organisation divide). However, opinion leaders and/or boundary spanners who hold an indifferent or negative opinion of the intervention may undermine it (intentionally or unintentionally), while an enthusiastic champion is likely to ensure core tasks are delivered and amplify enthusiasm, albeit on a smaller scale than the opinion leader or boundary spanner. Consequently, genuine support for the intervention appears to be a more important primary characteristic than influentiality or connections. This is hard to ascertain up front and is dependent on local cost/benefit judgements, but managers are well placed to identify likely candidates and, in combination with intervention staff, encourage increased appreciation of the intervention’s potential. Others have reported success in gaining support from people who were initially opposed to an intervention. [495] Alternatively, agencies might issue an internal call for LP candidates assuming that self-nominees are likely to be committed to the intervention and the work required to promote it.

→ Proposition 2: The LP should have credibility in relation to intervention goals.

Colleagues judged the suitability and effectiveness of their LP in relation to their credibility as an informed advocate for the intervention. Credible LPs had a professional reputation that aligned with the intervention goals (e.g. they modelled and espoused research-informed work practices). This point and the one above accord with the literature which indicates that in order for a colleague’s espousal to be meaningful they must be perceived as someone who believes in what they are saying and knows what they are talking about. [360, 518]

→ Proposition 3: The LP should have sound cross-organisational knowledge and connections

The intervention was more tailored, more creatively integrated, and better attuned to professional development expectations when LPs consulted with colleagues and shared their knowledge about organisational priorities, processes and learning norms with intervention designers. LPs’ ability to act as intervention intermediaries in this regard required them to have (or be able to rapidly acquire) a good understanding of their organisation and the people who work in it. This requires breadth: without boundary spanning skills, the efforts of champions may be restricted to highly localised contexts. [361] But it also requires depth: an ability to understand diverse perspectives and needs arising from complex contextual interactions, and to respond accordingly. [328, 502]

→ Proposition 4: The LP should have good interpersonal skills. Ideally, they will be friendly, approachable and well-liked.

Unsurprisingly, our data support assertions in the literature that people are more inclined to do things for people they like. This reminds us that ‘reach’ is about more than access. The quality of connections was just
as important as the quantity for supporting organisational understanding and engagement, including identifying and resolving concerns during implementation. The need for communication and project management skills is a given.

Managerial support

→ Proposition 5: Organisational leaders need to visibly back the LP as well as the intervention.

Strong, visible support for the intervention from managers was key in assuring LPs that their efforts—even when they verged on “nagging”—were seen as reasonable and warranted [see also 362]. Colleagues confirmed that strong support from above increased the LP’s authority and demonstrated they were acting on behalf of management. Others note that managerial support should encourage LP’s autonomy as overly specifying their approach could stifle enthusiasm and creativity. [494]

→ Proposition 6: If possible, the LP function should be incentivised within the organisation.

Enthusiasm for the intervention appeared to be enhanced when mechanisms or opportunities associated with the LP function benefited the LP professionally. This included formal professional development recognition (e.g. building the work into performance indicators); increased organisational exposure, status or connections; or furthering the LP’s own work. In most cases, this will be effective only if there is some congruence between the intervention goals and the LP’s day-to-day work. Protected time for the LP tasks to be conducted during work hours should be agreed; this might include time for LPs in multisite interventions to share ideas and experiences with each other. [519] A caveat: incentives should in no way pressure LPs to coerce participation.

Intervention team responsibilities

It is hard to overemphasise the importance of the relationships between the intervention team, the LP and organisational leaders. With the benefit of hindsight, these relationships would have been given a higher priority in our study.

→ Proposition 7: Intervention staff should provide CEOs, LPs and the LPs’ line managers with clear and realistic guidance about the attributes and demands of the LP function.

The strikingly different conceptualisations of the LP role indicate that, at a minimum, we must emphasise that LPs are skilled facilitators rather than administrators per se. Propositions 1-4 above provide the key messages for this exchange. The intervention team must describe the full scope of responsibilities and err in favour of over-estimating likely time commitments.

→ Proposition 8: Agencies should be supported to enact the role of LP flexibly where it does not compromise implementation fidelity.

Our findings indicate the benefits of a flexible approach in which core objectives and tasks are specified but the strategies for achieving them can be developed locally. [503] For example, agencies might prefer to divide the LP function between two members of staff with one taking responsibility for administrative tasks and another for creative input, persuasive communication, and higher level decisions. This has been effective in other studies, especially when those staff work (and are therefore likely to have influence) at different levels of the organisation. [519]
Proposition 9: Intervention staff should actively engage the LP in planning and problem-solving, treating them as a partner in the intervention rather than a conduit.

Where LPs shared detailed insider knowledge, employed creative strategies, and made suggestions for increasing the benefits of SPIRIT in their organisations, intervention activities were assessed by implementation staff and participants as more useful. This indicates that working with LPs as an intervention development partner, rather than as an implementer, would increase our ability to learn about and respond appropriately to local conditions, enhancing the relevance and fit of the intervention’s goals and activities. [493] ‘Ownership’ approaches have been highly successful in effecting and sustaining change. [1, 520] They enable interventions to focus less on diffusing knowledge and more on contributing to how it is shaped and applied. [521] LPs who co-owned the intervention would be more likely to understand it fully (genuine dialogue bypasses much of the formal communication that SPIRIT struggled with), believe in its potential benefits and be perceived by colleagues as authentic advocates. [493] However, this would depend on a fundamental philosophical agreement between the LP and the researchers about the goals and functions of the intervention. Such collaboration would also require a significant time commitment.

In a subsequent paper, we plan to examine the concordance between the LP attributes, perceptions and behaviours reported here (findings which are blinded to the quantitative results) and the observed intervention effects. We recognise that many factors will affect how SPIRIT was received, but believe that the propositions outlined above, together with our analysis of the centrality of each organisations’ research culture and trajectory of change, will help explain the trial outcomes.

Limitations

A limitation regarding the interviews was our inability to reach everyone who acted as an LP in their agency, and to interview as many senior managers as we would have liked, including the agency CEOs. Hence we may have missed some important perspectives. We were also unable to test our propositions formally. Thus, while they are sound representations of our findings across the six intervention sites, we do not know to what extent they provide useful applied guidance in identifying and working with LPs, nor how applicable they are to different organisational contexts.

Conclusions

This paper shows that the liaison people (LPs) who acted as facilitators of the SPIRIT study had a profound impact on how the intervention was implemented. LPs made informal cost/benefit analyses in which they weighed the value of participation against its demands and potential risks. Their different conclusions—influenced by their organisation’s mission, research utilisation norms, epistemological stance and leadership support—led to substantial variation in how they facilitated, promoted and tailored the intervention. This impacted on participation and engagement with the study across their respective organisations. LPs’ judgements about SPIRIT may have had a degree of self-fulfilment (they got what they put in), however, in some cases the intervention’s form and content may have been unsuitable for adaptations that could best address the organisations’ most pressing knowledge-to-practice needs. This indicates that the design of research utilisation interventions in policy agencies should incorporate potential participants’ views about the role of evidence in policy making and how local practices can be best supported. Nine propositions were developed from the data that may assist in identifying and supporting facilitator roles in interventions similar to SPIRIT and, potentially, inform a framework for evaluating attributes and conditions.
Chapter Nine: Participants’ experience of SPIRIT

“Realist evaluation allows us to learn from what works and why, offering lessons on how to effect change in settings that are complex, messy, unexpected, political—and decidedly human”

Adams et al. [522:274]

Overview

This study examines the question of how and why SPIRIT had the process effects it did. It addresses two overarching questions: How did people perceive SPIRIT? And with what process effects? As described in Chapter Four, the term process effect is used to describe a proximal intervention outcome such as whether people participate and how, their views of the intervention and the wider trial (e.g. data collection, the study’s communications), and their immediate and short-term responses to the intervention, including any changes. Others use the term formative outcomes. [13]

The analysis synthesises information from all the data sets over the 30-month trial duration, identifying interactions between the intervention as it was delivered and the many policymakers who participated in aspects of SPIRIT in the six policy organisations. The findings show substantial variation between the agencies and, importantly, they provide plausible tentative explanations for this variation that have implications for intervention and implementation improvement. The realist approach maximises the transferability of findings that, we hope, will be useful for others in designing, implementing and evaluating similar interventions where they want to maximise engagement and participation.

Unlike the ‘liaison person’ study in the previous chapter, participants were not invited to comment on the draft findings. This was because a far larger dataset had been synthesised and abstracted to such an extent that people would be unlikely to recognise themselves or their experiences as the liaison people had done. [523] Also, the analysis was conducted at a later date so many of the key participants had moved on, and the analytical approach is unusual so iterative dialogue would have probably been required to make sense of it. Overall, it was likely to make too many demands for dubious gains.
Abstract

Background: An intervention’s success depends on how participants interact with it in local settings. Process evaluation examines these interactions, indicating why an intervention was or was not effective, and how it (and similar interventions) can be improved for better contextual fit. This is particularly important for innovative trials like SPIRIT (Supporting Policy In health with Research: an Intervention Trial) where causal mechanisms are poorly understood. SPIRIT was testing a multi-component intervention designed to increase the capacity of health policymakers to use research.

Methods: Our mixed-methods process evaluation sought to explain variation in observed process effects across the six agencies that participated in SPIRIT. Data collection included observations of intervention workshops (n=59), purposively sampled interviews (n=76), and participant feedback forms (n=553). Using a realist approach, data was coded for context-mechanism-process effect configurations (retroductive analysis) by two authors.

Results: Intervention workshops were very well received. There was greater variation of views regarding other aspects of SPIRIT such as data collection, communication, and the intervention’s overall value. We identified nine inter-related mechanisms that were crucial for engaging participants in these policy settings: 1. Accepting the premise (agreeing with the study’s assumptions), 2. Self-determination (participative choice), 3. The Value Proposition (seeing potential gain), 4. “Getting good stuff” (identifying useful ideas, resources or connections), 5. Self-efficacy (believing “we can do this!”), 6. Respect (feeling that SPIRIT understands and values one’s work), 7. Confidence (believing in the study’s integrity and validity), 8. Persuasive Leadership (authentic and compelling advocacy from leaders) and 9. Strategic Insider Facilitation (local translation and mediation). These findings were used to develop tentative explanatory propositions and to revise the program theory.

Conclusion: This paper describes how SPIRIT functioned in six policy agencies, including why strategies that worked well in one site were less effective in others. Findings indicate a complex interaction between participants’ perception of the intervention, shifting contextual factors, and the form that the intervention took in each site. Our propositions provide transferable lessons about contextualised areas of strength and weakness that may be useful for maximising engagement and participation in similar studies.

27 The published version of this manuscript is in Appendix 13.
Background

This paper presents a realist analysis of how a novel, multi-component intervention trial designed to increase research use capacity known as SPIRIT (*Supporting Policy In health with Research: an Intervention Trial*) functioned in six health policy agencies. Data from a mixed-methods process evaluation is used to unpack the processes of engagement and participation that were hypothesised to mediate the intervention’s success. These intermediate impacts are conceptualised as process effects (see Box 1 for definitions). We do this by describing what was delivered in the intervention and what process effects were observed, then identify explanatory Context+Mechanism→Process effect configurations that show: how the intervention and the trial more broadly was perceived by participants, why this varied across the participating organisations, and how these perceptions affected receptivity to the intervention’s ideas and resources. A realist approach is used because it supports rigorous comparative analysis of how those targeted by an intervention make sense of what it offers, and how this is shaped by context. [15, 23, 524]

Understanding interventions

Interventions—planned activities to change individual, group and/or organisational behaviour—are not passively received, but are actively shaped by the people who participate in them and the circumstances in which they are delivered. [5, 12, 16] Understanding the ways in which participants interact with and perceive an intervention is vital for determining how and why it was, or was not, effective. [17] This requires moving beyond measures of participant satisfaction—sometimes derided as “happy face evaluation” [525]—towards methods which delve into “the complexity, flux and contextual variation that inevitably occurs in real life situations.” [25]

Many organisational capacity building interventions fail because they do not take sufficient account of participants’ workplaces. [526] Successful interventions introduce strategies (ideas, activities and resources) that are contextually apt [17, 477] and which are therefore able to produce desired interactions. [23] For example, in organisational interventions, participants’ perceptions and interactions are affected by factors such as the organisation’s culture [258], its history of change [392, 527], staff heterogeneity [528] and trust in management. [527]

Information about how implementation interacts with people and place over the course of an intervention is frequently overlooked [7]; yet it is necessary for making informed assessments about the worth, adaptability and transferability of strategies designed to bring about individual or organisational change. [25] In multi-component interventions it is often impossible to disentangle which components were more or less effective, or what variations in combination might maximise effectiveness. [529] These interventions frequently trigger unanticipated causal processes and have unpredictable impacts that standardised measures are unlikely to capture. [405] This may be especially important for interventions where participants have involvement in the tailoring and/or delivery of an intervention, since their attitudes towards its content, form and goals are likely to have profound impacts on what is delivered and how it is received. [265, 530] Indeed, there is an established link between outcomes and the ways that participants gauge the quality of their involvement in tailoring the scope, content and process of flexible interventions. [16]

Context-sensitive design, implementation and evaluation are particularly pressing for interventions that attempt to increase the use of research in policy processes. Policymaking is “a contested arena of negotiation... messy, complex, and serendipitous” [35] where research—and researchers [94]—are used strategically. [33, 34] Macro-level political and institutional factors influence how policymakers and policy organisations engage with and make use of research [33], and will therefore mediate their relationships with research utilisation interventions. Given that the use of research is cultural and rhetorical as well as technical [36], where an intervention promotes greater use of research, or claims to be evidence-based, participants
may actively critique that premise. [37, 38] Thus, determining if and how such an intervention is compatible with participants’ beliefs and practice norms is critical.

Despite these arguments, many interventions are reported (and, by implication, conducted) with minimal consideration of the interactions between the intervention activities, the people who took part, and the circumstances that mediated this relationship. [25, 531] As Clark and colleagues note, “Little research has explored individuals’ experiences of programmes or examined how programme dimensions lead to changes in behaviour. ...individuals’ meanings, experiences and reactions to the programme and the effects of their wider context are simultaneously disregarded”. [532:518] Realist process evaluation is well equipped to redress these oversights. [15, 23]

### Box 1. Definitions of key concepts used in this paper

<table>
<thead>
<tr>
<th>Concept</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Context</strong></td>
<td>In realist terms, context are any systems, structures or conditions that affect outcomes. This includes individuals’ attributes and social interactions. [23]</td>
</tr>
<tr>
<td><strong>Mechanism</strong></td>
<td>Mechanisms are what makes an intervention work: “They are not the observable machinery of program activities, but the response that interaction with a program activity or resource triggers (or does not trigger) in the reasoning and behaviour of participants”. [533]</td>
</tr>
<tr>
<td><strong>Process effects</strong></td>
<td>These are proximal impacts that influence intervention outcomes or are of evaluative interest for other reasons (e.g. they help explain unexpected variation in implementation). Others use the term formative outcomes. [13] Desired process effects are those the investigators consider to be prerequisites for a successful intervention.</td>
</tr>
<tr>
<td><strong>Program theory</strong></td>
<td>This is, “An explicit theory or model of how an intervention contributes to a set of specific outcomes through a series of intermediate results.” [398] Program theory should be plausible, useful and consistent with the evidence.</td>
</tr>
<tr>
<td><strong>Proposition</strong></td>
<td>Propositions are generalised theoretical statements grounded in the data. [505] In realist evaluation they link and condense information about contexts, mechanisms and outcomes. Propositions are refined through empirical testing but remain fallible. [534]</td>
</tr>
<tr>
<td><strong>Realist process evaluation</strong></td>
<td>Process evaluation helps explain how an intervention had its effects. [17] Realist process evaluation applies realist principles to this process. It investigates causal patterns (known as demi-regularities) to show how intervention strategies may be operating under what conditions to generate process effects for which groups. [23]</td>
</tr>
<tr>
<td><strong>Retroduction</strong></td>
<td>A form of analysis that attempts to explain phenomena by theorising about what mechanisms are capable of producing them. [535] This “involves constant shuttling between theory and empirical data, using both inductive and deductive reasoning.” [536]</td>
</tr>
</tbody>
</table>

**The study being evaluated: SPIRIT (Supporting Policy In health with Research: an Intervention Trial)**

SPIRIT was a stepped wedge cluster randomised trial that tested the effects of a novel intervention designed to increase the capacity of health policy agencies to use research. Six organisations in Sydney, Australia participated. Five were state government agencies and one was a national organisation funded by the federal government. An agency was eligible to participate if (a) a significant proportion of its work was in health policy or program development, and (b) there were at least 20 staff involved in health policy, program development or evaluation. A sampling frame was drawn from Government websites that listed all New South Wales and Australian government health policy and program agencies located in Sydney. Members of the investigator team reduced this list to 16 potentially eligible agencies and ranked as highest those with the greatest focus on health and the largest numbers of relevant staff. The top six agencies were invited to take part, and all agreed. [227] Each agency’s CEO signed an organisational-level agreement to participate in SPIRIT and nominated a liaison person: an internal member of staff who would be responsible for coordinating SPIRIT in their setting for the duration of the trial. There were six rounds of outcome data collection using three evaluation tools. These are described in detail elsewhere. [227, 376-379]
The intervention aimed to increase agency capacity to use research in relation to three goals: 1. The organisation and staff value research more, 2. More tools and systems are in place to support research engagement actions and the use of research, and 3. Staff have greater knowledge and skill in research engagement actions and the use of research. SPIRIT’s design was informed by an action framework [375] and underpinning change principles that reflected composite theory from psychology, organisational science, adult learning and the research utilisation literature. [227] The intervention comprised multiple components hinging on interactive workshops such as research skills seminars, exchange forums with researchers, and a leadership program targeting senior managers. Other activities included the provision of tools and resources (such as an online research portal), practice using systems for commissioning research reviews, analyses or evaluations, and CEO espousal of research informed policymaking (Figure 9.1). Agencies could choose options within and tailor many of the components to address local priorities. Each agency was asked to identify two lists of potential participants: 1. all staff involved in policy or program development, implementation or evaluation who would be invited to take part in intervention activities and data collection; and 2. managers who would take part in the leadership program and promote SPIRIT.

**Figure 9.1. SPIRIT intervention model**

![SPIRIT intervention model diagram]

Key: Ovals contain intervention principles, white boxes show intervention targets and strategies, and the shaded box shows agency-based implementation supports.

An onsite introductory information session preceded the intervention and data collection in each site. The round of data collection that took place immediately before the intervention functioned as an audit and was followed by a feedback forum in which the lead investigator facilitated a deliberative dialogue with leaders about their agency’s findings. Intervention goals targeting research engagement and use were identified during this process. Agency leaders considered how they would like to use SPIRIT’s options to address these goals and, if applicable, any additional (non-SPIRIT) strategies for reaching their goals.

External research and policy experts were contracted to deliver workshops. They were briefed on SPIRIT’s ‘change principles’ and their workshop’s objectives. The content of the tailored workshops was negotiated with the agency’s liaison person, with input from presenters. Members of the SPIRIT research team coordinated the development and delivery of workshops and other intervention activities. Each site had a
dedicated knowledge broker from the SPIRIT team who acted as the onsite ‘face’ of SPIRIT, negotiated tailoring, and attended all intervention activities. An in-depth, mixed methods process evaluation informed by realist thinking was conducted in parallel with the intervention. This paper is based on that data.

The role of process evaluation

Process evaluation investigates an intervention's implementation, change mechanisms and contextual interactions in order to explain (insofar as this as possible) how and why the intervention functioned as it did in each intervention site. [405] Process evaluation does not determine whether study outcomes are achieved, but it can identify process effects - proximal impacts of an intervention that make achieving outcomes more or less likely. [537]

Aims

Using a realist evaluation approach [15, 23, 52, 538] we aimed to generate transferable learning in relation to two questions: 1. To what extent did SPIRIT achieve the desired process effects in each agency? and 2. How were these process effects generated, i.e. what mechanisms seem to account best for the patterns of engagement and participation observed across all agencies?

Methods

Realist evaluation

The SPIRIT process evaluation comprised a fidelity assessment and a theory-driven exploration of the interaction between the intervention, participants and the implementation circumstances, with the expectation that this would probably take a different form in each of the six agencies. [384] Theory-driven evaluation seeks to uncover causal pathways [539] and is well suited for understanding how multicomponent interventions function in complex real world settings. [540] In this study we adopt a particular theory-driven approach—realist evaluation [3]— following the methods associated with Pawson [15], Pawson and Tilley [23], and others in the RAMESES II project. [52] Realist evaluation focuses on an intervention’s underlying theory as its unit of analysis [15, 23], with the aim of determining “what works, for whom, in what circumstances, and how”. [23, 53] Realists posit that interventions introduce ideas and opportunities that generate effects in conjunction with participants’ reasoning and resources. Thus the interaction between intervention activities and the contexts of each intervention site will determine what (if any) mechanisms are activated, and what outcomes (intended and unintended) are generated. [541, 542]

We used a realist approach because it maximises the transferability of findings and operational learning from one setting to another (an enduring concern in intervention evaluation [188]), while also recognising complexity and the need to look beyond one-size-fits-all ways of responding to problems [15, 23, 543, 544]. Realist evaluation has been used effectively in studies of policy processes [545], implementation research [522], knowledge exchange [233] and evaluations of flexible intervention trials [265, 532]: making it especially suitable for addressing the methodological challenges presented by a multi-component, novel and theoretically eclectic trial like SPIRIT (outlined in detail elsewhere [503]).

Importantly, analyses arising from realist evaluations are tentative, claiming only to be an informed hypothesis of “how something might be” [44] rather than a definitive version of reality. These hypotheses accrue plausibility when tested in further studies, but remain open to revision or rejection if alternative theories are more convincing. [541] In our study, data collection, management and analysis were concurrent, thus we were continually testing and revising hypotheses within and across the six intervention sites over the 30-month study, but our findings are embryonic in realist terms.
Initial program theory

Realist evaluation develops, tests and refines program theory. SPIRIT was informed by a mixture of formal theory and experiential knowledge [227], and had both a well-articulated action framework [375] and clear principles about what should be provided [503], but did not offer hypotheses about the mechanisms that would generate increased capacity to use research. Based on existing trial materials and discussions with the designers we articulated the overarching program theory to make the intended causal pathway more explicit so that we could critique the assumptions underpinning the intervention design. [15, 23, 546] This was refined and agreed through further consultation:

SPIRIT will engage and motivate agency leaders to ‘own’ the intervention using audit feedback, deliberative goal-setting and program tailoring. This agency-driven approach will generate a priority-focused program that offers locally relevant practice support and accommodates differences in agencies’ values, goals, resources and remits. The program will comprise a suite of andragogical activities, tools, and connection across the research-policy divide that provide resources and build knowledge, skills and relationships. It will be supported via modelling and opinion leadership by agency leaders and dynamic external experts. CEOs will promote SPIRIT in their agencies and liaison people will facilitate the tailoring and implementation. These strategies will act synergistically to stimulate and resource participants at different organisational levels, leading to changes in values, practice behaviours and agency processes. This will facilitate increased use of research in policy processes.

This pathway informed the data collection, providing pointers about what to look for, but was used flexibly (rather than as a rigid investigative framework) as befits an exploratory study. We also looked for unintended effects, and considered alternative causal pathways that might better explain observed effects. The data offered the opportunity to develop a much richer understanding of the social processes and interactions than had previously been possible.

Process effects

The program theory was used to identify desired process effects via discussion with the study designers. We then explored how these process effects were achieved in each setting for the range of targeted participants, or why they were not. Our conceptual framework for this work was informed by the implementation science literature that focuses on social processes and interaction in interventions (e.g. [12, 13, 37, 350, 547-549]).

Data collection

Causation, and the mechanisms that generate it, are seldom observable. [23] So in realist evaluation data is triangulated to identify the interactive patterns that can most plausibly explain how the intervention led to the observed outcomes. [550] Quantitative data is helpful for identifying outcomes [15] while qualitative methods are usually necessary “to discover actors’ reasoning and circumstances in specific contexts”. [46] We used the following methods to capture information:

1. Semi-structured interviews with 5-9 participants from each agency early in the intervention period (n=33) and post-intervention (n=43). Interviewees were purposively selected for maximum variation in work roles, attitudes to research, and experiences of SPIRIT in order to explore the breadth of dimensions expected to influence interactions with the intervention. [17] Open-ended questions and prompts explored interviewees’ work practices and contexts, and their experiences and perceptions of SPIRIT, including their causal explanations for any change. The interview questions are available elsewhere. [384] This combination of context-, causal- and impact-focused questions across diverse participants was used to refine theory about what was working (or not) for whom and in what circumstances.

2. Observations of intervention workshops (n=59), and informal opportunistic conversations with participants before and after workshops. Workshops were audio recorded and fieldnotes were written immediately
afterwards. A checklist was used for fidelity coding through which we monitored the extent to which ‘essential elements’ of the intervention were delivered [detailed elsewhere 503].

3. *Anonymous participant feedback forms* (n=553). These comprised Yes/No ratings on six statements: (i) The workshop was interesting, (ii) The workshop was relevant to my work, (iii) The workshop was realistic about the challenges and constraints of our work, (iv) The presenter had appropriate knowledge and skills, (v) It is likely that I will use information from this workshop in my work, (vi) It is likely that SPIRIT will benefit my agency (Additional file 1). Some workshops had additional items, e.g., the forms for audit feedback forums included items about the clarity of the data and participants’ confidence that SPIRIT would be adequately tailored for their agency. All forms contained three open-ended questions: 1. ‘What worked well?’, 2. ‘What could be improved?’, and 3. ‘Any other comments?’ Forms were distributed prior to intervention workshops and completed immediately afterwards.

4. *Formal and informal interviews with the people implementing SPIRIT and the commissioned presenters.*

5. *Limited access to information from the interviews conducted as part of SPIRIT’s outcome evaluation.* These interviews focused on: (a) organisational support for research use (n=6), and (b) the role of research in the development of a recent policy or program (n=24). We reviewed transcripts from the first round of interviews (prior to the intervention), but thereafter were blinded to this data so that it would not influence the ongoing process evaluation analysis.

**Data management and analysis**

**Qualitative data**

Data was initially analysed for the whole process evaluation. Interview data was managed using framework analysis [406] within NVivo v.10 [504] and used to develop descriptive case studies [551] in combination with data from the fidelity assessment, running memos for each agency, interviewee memos, and thematically coded data from field notes and the open-ended questions in feedback forms. These case studies described: (a) each agency’s context, change trajectory, workforce and practice norms, (b) their research use practices and culture, (c) how SPIRIT was implemented in each setting, and (d) the interactions between (a), (b) and (c). Framework categories and the structure of the case studies were iteratively developed from *a priori* concerns (such as the constructs the intervention was targeting and the hypothesised causal pathway), and from themes that we identified using inductive analysis. [552, 553] The method of constant comparison [554] was used to query and refine the initial program theory and other emergent hypotheses throughout the trial. This work is described in more detail elsewhere. [384]

**Quantitative data**

For each agency, we calculated the number and percentage of feedback forms responding ‘YES’ to each of the six statements outline earlier. In calculating these frequencies, the four different types of workshops (symposia, research exchanges, leaders’ forums, and audit feedback forums) were aggregated.

**Realist analysis**

Using the data described above, we explored the hypothesised pathway identified in the program theory, and sought to identify any other pathways leading to the interventions’ observed process effects. This included consideration of unanticipated impacts reported by participants or members of the implementation team. [540] We used the observed process effects as a starting point: although realist evaluation depicts outcomes (or, in our case, process effects) as the final step in the sequence, the analysis tends to start with identifying effects, then working backwards to investigate the conditions (context and mechanisms) that caused them. [555] We traced connections to and from observed process effects asking *What caused this?, Why didn’t this unfold as anticipated?* and *What best explains these different responses between agencies?*

We employed a *retroductive* analytical approach that attempts to explain phenomena by theorising about what mechanisms are capable of producing them. [535] This involves studying events “*with respect to what*
may have, must have, or could have caused them. In short it means asking why events have happened in the way they did". [Olsen and Morgan in 541] In accordance with realist evaluation principles, we focused on the interaction of SPIRIT with features of each agency’s context that appeared most likely to have influenced process effects [540] [533], and developed explanatory configurations of the patterns we saw in the data. In realist evaluation these are typically called Context+Mechanism—Outcome (CMO) configurations [15, 23] but, because the ‘outcomes’ of interest in process evaluation are process effects rather than study outcomes, we have called them Context+Mechanism—Process effect configurations here. Propositions were then developed to summarise each configuration. This work depended on using each type of data to query, explain and balance the other to reach accounts of what happened and why that were as comprehensive as possible. [109, 556] Original data sources were revisited frequently.

This analysis relied on iterative cycling between data and likely explanations and so incorporated both inductive and deductive reasoning. [557] We looked for evidence of causal mechanisms, and for evidence that supported, discounted or nuanced current causal hypotheses both in real time (as the intervention unfolded) and retrospectively (reviewing data already collected). Throughout this process we sought to identify where our evolving Context+Mechanism—Process effect configurations aligned with existing theory: we revisited the theories used to inform the development of SPIRIT, asking to what extent did these theories support the patterns we were observing in the data, and also considered other theories that might better explain our findings. This was not a comprehensive search for theory, but a pragmatic exploration of possible explanations from bodies of literature that we felt were most likely to offer insights. Where theories did not fit with the data across all agencies, they were rejected. See Appendix 14 for an overview of the theory that we found to be most useful.

AH, who led the process evaluation, reviewed and coded all data sources. SB, who contributed to the process evaluation design and analysis throughout the trial, independently reviewed a proportion of interview transcripts and fieldwork memos. Our preliminary Context+Mechanism—Process effect configurations overlapped extensively and were workshopped with further reference to the wider data set to develop agreed configurations. This was not a straightforward process: it took months of emersion in the data, argument about what we were seeing and how this should be conceptualised to reach the point where we both felt the findings accurately represented the data and were appropriately theorised. During this process I realised that the variation I was seeing between agencies required a more nuanced conceptualisation of mechanisms than we had been using, i.e. there was evidence of mechanisms being activated, but also of mechanisms being weak or entirely absent in some sites28. This led to us describe ‘how mechanisms functioned’ in the results – a more detailed and narrative account than is usual in CMO tables. Further discussion with our co-authors (most of whom had firsthand experienced of the trial) further refined the findings.

---

28 This was aided by Dalkin and colleagues’ assertion that mechanisms may vary in intensity rather than simply being present or absent. See Dalkin, S.M., et al., What’s in a mechanism? Development of a key concept in realist evaluation. Implementation Science, 2015. 10:49.
Results

In this section we describe the implementation of the SPIRIT intervention, outline the observed process effects, and then attempt to explain how these effects were generated using Context+Mechanism→Process effect configurations. Lastly, we present the revised program theory.

Implementation

As Table 9.1 shows, some aspects of SPIRIT were delivered with a high degree of implementation fidelity: every agency received audit feedback and the intended number of components on the topics they requested. Intra-organisational processes that were outside the control of the implementation team had greater variation. The promotion of SPIRIT and much of its administration depended on the attitudes and behaviours of liaison people and each organisation’s leaders, and to a lesser extent, the expert presenters commissioned for each workshop. This resulted in some loss of SPIRIT’s theoretical fidelity, i.e. the extent to which the intervention delivered its ‘essential elements’ (these are discussed in more detail elsewhere [503]).

For example, the essential elements stipulated that workshops should be non-didactic so the presenters should encourage participants to contribute as much as possible. Many workshops were highly interactive, such as the deliberative audit feedback forums, but others were not. This was because: 1. The expert presenters sometimes overrode their briefing to facilitate discussion; 2. Liaison people occasionally tried to maximise value by cramming content into workshops which limited opportunities for participation; and 3. Unexpectedly, the agencies seldom took up offers to co-design and co-present workshops.

In some sites, SPIRIT’s reach was constrained more than anticipated. Agency 6, for example, chose to focus some components of the intervention on one group of staff and limited participation accordingly. In Agency 3, managers attempted to minimise the onerousness of data collection by excluding some eligible staff from invitations to complete surveys. Agencies also defined their leadership groups quite differently, resulting in wide variation in the numbers and organisational roles of participants in the leaders’ program.

Process effects

Table 9.2 describes SPIRIT’s process effects, i.e. the actions, behaviours and responses hypothesised to be necessary for SPIRIT to generate the capacity-related outcomes measured in the trial. Column 1 lists the process effects both for the intervention and the trial evaluation; we include the latter because of their impact on the quality of the evaluation and the way that SPIRIT as a whole was perceived. Column 2 presents a summary of our observations about the extent to which these process effects occurred. Column 3 shows the data sources for our observations.
Table 9.1. Summary of SPIRIT intervention implementation fidelity

<table>
<thead>
<tr>
<th>Fidelity items</th>
<th>Agency 1</th>
<th>Agency 2</th>
<th>Agency 3</th>
<th>Agency 4</th>
<th>Agency 5</th>
<th>Agency 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEO emails espousing the value of research</td>
<td>All sent by CEO</td>
<td>Some sent by liaison person</td>
<td>All sent by liaison person</td>
<td>Some sent by liaison person</td>
<td>All sent by CEO</td>
<td>All sent by CEO</td>
</tr>
<tr>
<td>Use of systems for commissioning research</td>
<td>✓ Rapid review</td>
<td>✓ Evaluation plan</td>
<td>✓ Rapid review</td>
<td>✓ Linked data analysis</td>
<td>✓ Evaluation plan</td>
<td>✓ Rapid review</td>
</tr>
<tr>
<td>Provision of tools and resources</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Essential elements* delivered in workshops:</td>
<td><strong>Content</strong></td>
<td><strong>Interaction</strong></td>
<td><strong>Content</strong></td>
<td><strong>Interaction</strong></td>
<td><strong>Content</strong></td>
<td><strong>Interaction</strong></td>
</tr>
<tr>
<td>- Introductory session</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Extensive</td>
</tr>
<tr>
<td>- Audit &amp; feedback 1</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Extensive</td>
</tr>
<tr>
<td>- Audit &amp; feedback 2</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Moderate</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Extensive</td>
</tr>
<tr>
<td>- Skills seminar 1</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Extensive</td>
<td>Moderate</td>
<td>Extensive</td>
</tr>
<tr>
<td>- Skills seminar 2</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Extensive</td>
</tr>
<tr>
<td>- Skills seminar 3</td>
<td>Extensive</td>
<td>Limited</td>
<td>Limited</td>
<td>Extensive</td>
<td>Moderate</td>
<td>Limited</td>
</tr>
<tr>
<td>- Leaders’ program 1</td>
<td>Moderate</td>
<td>Limited</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Unknown</td>
</tr>
<tr>
<td>- Leaders’ program 2</td>
<td>Extensive</td>
<td>Moderate</td>
<td>Limited</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Extreme</td>
</tr>
<tr>
<td>- Exchange forum 1</td>
<td>Extensive</td>
<td>Moderate</td>
<td>Limited</td>
<td>Extensive</td>
<td>Extensive</td>
<td>Extreme</td>
</tr>
<tr>
<td>- Exchange forum 2</td>
<td>Extensive</td>
<td>Moderate</td>
<td>Extensive</td>
<td>Moderate</td>
<td>N/A: e-bulletin chosen</td>
<td>Moderate</td>
</tr>
<tr>
<td>- Exchange forum 3</td>
<td>Extensive</td>
<td>N/A: e-bulletin chosen</td>
<td>N/A: e-bulletin chosen</td>
<td>N/A: e-bulletin chosen</td>
<td>N/A: e-bulletin chosen</td>
<td>N/A: e-bulletin chosen</td>
</tr>
<tr>
<td>Total number of activities delivered</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
</tbody>
</table>

*The essential elements for each workshop are clustered in two categories: **Content** includes the core topics, activities, messages and resources agreed with the agency; while **Interaction** includes how attendees participated in activities and, crucially, the extent to which they contributed to the content via discussion. Details about the methods and conduct of the fidelity assessment are provided elsewhere. [503]
Table 9.2. Overview of SPIRIT’s process effects and data sources

<table>
<thead>
<tr>
<th>Desired process effects for the trial</th>
<th>Observed process effects</th>
<th>Supporting data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Leaders espouse SPIRIT and its goals</td>
<td>All CEOs disseminated initial information about their agency’s participation in SPIRIT, but only four had a continuing visible role in supporting the intervention, e.g. sending updates and attending workshops. Some executive members participated in each site, but to very different extents ranging from a ½ hour ‘drop in’ to repeated and enthusiastic participation. Many managers talked about SPIRIT in team meetings and encouraged their staff to attend.</td>
<td>Interviews at two time points (early-intervention ‘context’ and post-intervention ‘perceptions and impact’), ad hoc conversations with participants</td>
</tr>
<tr>
<td>2. Liaison people facilitate the intervention effectively</td>
<td>The use of a liaison person was very effective in the sites where the liaison person was enthusiastic about SPIRIT. Four of the six worked hard to promote, tailor and administer the intervention, harnessing insider knowledge and using creative strategies. The other two did not tailor or promote the intervention as thoroughly and expressed negative views to colleagues about SPIRIT.</td>
<td>Observations of workshops, interviews and conversations as above, feedback from the SPIRIT team about their communications with liaison people</td>
</tr>
<tr>
<td>3. Targeted policymakers participate in, and are receptive to, intervention activities</td>
<td>Participation levels were good in that they met the SPIRIT team’s expectations for each site. Each agency targeted different groups for different components so proportions and types of participants varied, but liaison people were satisfied with attendance and occasionally surprised by very high numbers. Attendance at workshops averaged between 11-20 participants per workshop, with between 102-158 total occasions of attendance across the six sites. There was full participation in other activities (e.g. trialling the commissioned research services). Receptivity varied tremendously within, but especially between, agencies: see next section for more details, including possible reasons.</td>
<td>Quantitative fidelity data from observations (using check lists and sign-in sheets), observations, interviews and conversations as above</td>
</tr>
<tr>
<td>4. Participants actively contribute to the content of those activities</td>
<td>Where there was opportunity, participants contributed greatly to workshop content via questions, discussion and case examples. Interactivity was limited on some occasions in all agencies, usually because the presenter provided few opportunities. In larger groups more senior staff tended to dominate, but other participants said this was still useful. Some liaison people helped craft workshop content and provided agency-based case examples. One agency co-presented a workshop. The agency staff nominated to test the research commissioning service were actively involved.</td>
<td>Observations of workshops, including descriptive accounts of interactions and dynamics</td>
</tr>
<tr>
<td>5. Participants identify potentially useful ideas, techniques and/or resources</td>
<td>94% of those who completed a feedback form said they found workshops to be both relevant to their work and realistic about policy challenges and constraints. Many interviewees identified specific benefits from SPIRIT, including improved awareness of useful researchers and research resources, understanding of the evidence relating to a policy problem, and access to existing agency resources.</td>
<td>Participant feedback forms, observations of workshops, interviews and ad hoc conversations with participants and liaison people</td>
</tr>
<tr>
<td>6. Participants use, or plan to use, these ideas, techniques and/or resources</td>
<td>Workshops facilitated less discussion than intended about how learning might be applied, but 95% of participants who completed a feedback form agreed, “It is likely that I will use information from this workshop in my work”. Some interviewees said they planned to use ideas or resources, and a few had done so, especially newer staff. Three liaison people had managerial-approved plans underway for research-focused education and/or systems improvement, e.g. mandated consideration of research in policy proposals. Two agencies had plans to use their commissioned research products.</td>
<td></td>
</tr>
<tr>
<td>Desired process effects for the evaluation</td>
<td>Observed process effects</td>
<td>Supporting data sources</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>--------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>7. Liaison people facilitate data collection effectively</td>
<td>All liaison people facilitated data collection sufficiently, although it was occasionally delayed and required prompting. Where liaison people championed SPIRIT they used additional strategies to encourage participation in data collection, in one agency this achieved a 100% response rate.</td>
<td>Outcome measures completion figures, interviews with participants and liaison people, feedback from SPIRIT team</td>
</tr>
<tr>
<td>8. Targeted participants take part in data collection</td>
<td>In all agencies, there was full participation in the two interview based measures, but more variable responses to the anonymous online survey. Response rates dipped in the second measurement point, but stabilised after the survey was shortened. Overall the online survey response rate was 56%. There was a mean 74% response rate for process evaluation feedback forms. Only ¾ of invitees took part in a process evaluation interview.</td>
<td>Outcome measures completion figures, interviews with participants and liaison people</td>
</tr>
<tr>
<td>9. The benefits of the intervention are judged to outweigh the burdens of the trial</td>
<td>Interviewees differed considerably in their assessments of the intervention, but where they felt it had benefits these were deemed to outweigh the trial’s burdens. This included those liaison people who championed SPIRIT from the start. Workshops with high profile and dynamic “service-orientated” presenters were especially valued. Nearly 98% of all feedback form respondents agreed with the statement, “It is likely that SPIRIT will benefit my agency”.</td>
<td>Early-intervention and post-intervention interviews, ad hoc conversations with participants and liaison people, feedback form data</td>
</tr>
</tbody>
</table>
How were these process effects generated?

We identified nine primary causal mechanisms, as shown in Figure 9.2.

The Context+Mechanism→Process effect configurations for each mechanism are presented in the following section. Each of the configurations begins with an overview of the context pertaining to that mechanism; a description of how we believe the mechanism functioned; how it generated process effects; and how process effects differed between participating agencies. A proposition that summarises the hypothesised causal pathway precedes each configuration.

Cross-references to other mechanisms are in shorthand so that mechanism 1 reads as M1, etc. Similarly, agency numbers are shortened so that Agency 1 is shown as A1, and so on. Inevitably, this is a highly truncated presentation of our findings. For those who seek more detail, a narrative description of the data that informed our identification of each mechanism can be found in Additional file 2. This additional file provides an ‘evidence link’ between the data and the findings that follow.
Targeted participants were busy, time pressured professionals in hierarchical organisations. Most were experienced in designing and implementing programs, but their research and evaluation skills and confidence varied widely. Agencies dealt with complex politicised problems across diverse sectors of health, often requiring systems change. ‘Evidence’ was multifaceted and contested, frequently driven by key stakeholders. Real-world implementability was king. Existing levels of research engagement varied considerably within and across agencies, as did the perceived usefulness of research and researchers. Internal resources tended to be underutilised. Each agency had established practices and professional development norms, and most were on a clear change trajectory. All were subject to government budgetary fluctuations and all had experienced recent organisational restructures or practice change initiatives.
## Mechanism 1. Accepting the premise

### Proposition:
Where participants regard the intervention’s form, goals and assumptions as compatible with their agency’s remit, values, practice norms, trajectory of change and current priorities—and providing there is perceived room for improvement—they accept the premise of the study and are receptive to what it offers.

### Context
Each agency had existing goals, values, resources, practices and change trajectories. They viewed their capacity to use research, and the importance of increasing this capacity, quite differently. There were diverse norms about what evidence is and how it should be developed, which were affected by their primary stakeholder groups. In the wider environment there was increasing emphasis on action-based research (partnering with practitioners to produce research fit for immediate decision-making). [558]

### How did mechanism 1 function?
Potential participants accepted or rejected the premise of SPIRIT based on:
1. The compatibility of SPIRIT's goals & assumptions with local conceptualisations of evidence and its role in policymaking, including how research related to the agency’s remit, values and practice norms; stakeholder relationships; and change trajectory.
2. The compatibility of SPIRIT’s form – whether its design was congruent with local conceptualisations of ‘good’ or ‘appropriate’ intervention/trial models.
3. Relative advantage: if participants believed they or their agency would benefit from increased use of research, i.e. they saw a need for SPIRIT.
4. Relative priority: they saw this need as immediate, i.e. there was some urgency.

Accepting the premise functioned on a continuum. Many policymakers expressed uncertainty rather than a firm view, and modified their view (usually becoming more accepting) while the intervention was underway. Participants’ conception of SPIRIT’s premise did not always align with the designers’ conception.

### How did mechanism 1 generate process effects?
When they accepted the premise of SPIRIT, leaders were more inclined to espouse the intervention and its goals. Liaison people facilitated the intervention more effectively, and targeted policymakers were more enthusiastic about participation and receptive to content.

Where policymakers dismissed SPIRIT’s premise they said they were unenthusiastic about participation (yet many did participate) and had low expectations of content. Liaison people who rejected the premise admitted they did not champion SPIRIT or facilitate the intervention as effectively as they could have, but they managed data collection satisfactorily.

Some leaders who had reservations about the premise expressed their doubts to staff, but also encouraged staff to participate in specific intervention activities.

### Agency comparisons
Most potential participants in A1, A5 and A6 saw SPIRIT as addressing “a real need” and were open to what it had to offer. A3 staff supported the premise but many felt it did not apply to them as they had “no room for improvement”. A few interviewees in all agencies dismissed the premise of SPIRIT, but especially in A2 and A4. However, many of these participated in at least one workshop, either because it was expected or because the potential merits of individual workshops (M3) overcame reservations about the intervention/trial as a whole.
Mechanism 2. Self-determination

Proposition: Where participants have scope and support to shape the form and function of the intervention, and where they make efforts to do so, they will (a) positively enhance the relevance and applicability of content and (b) invest in the intervention. Involvement in shaping even limited aspects of the intervention can deliver benefits. Participants also need to have real choices about participation. Self-determination is linked to M4, M5 and M6.

Context

Externally designed interventions often feel imposed. SPIRIT aimed to enable agencies to identify local goals and tailor workshop content, but agencies were time pressured. Participants had extensive expertise in drafting policy, and many were experienced program designers and implementers using increasingly collaborative, bottom-up models. They viewed SPIRIT through this professional lens. The intervention’s start date was randomised. CEOs decided if their agency would participate and nominated liaison people. Agencies encouraged participation but, as others note, participation in organisational interventions can feel ‘expected’. [511]

How did mechanism 2 function?

Self-determination is the feeling of having control. [559] Interventions that foster self-determination share power and allow participants to pursue a variety of goals. Participants felt they had some control in relation to SPIRIT when there was: 1. Flexibility – scope to use the intervention to address their needs. 2. Decision-making support the audit feedback and deliberative processes helped leaders to make informed decisions, and those involved in tailoring received guidance about what could be achieved and how to do it. 3. Locally shaped content – managers and liaison people actively tailored goals and content, and ensured colleagues had a say in it (M9); interactive workshops enabled participants to drive content; and participants co-designed and co-presented workshops. 4. Choice about whether to take part in the intervention and data collection or be a liaison person, irrespective of managerial expectations. Self-determination had to be tempered with judicious decision-making. Some choices backfired such as when liaison people crammed content into workshops which overwhelmed participants.

How did mechanism 2 generate process effects?

Tailoring and interactivity were consistently viewed as critical for getting value out of participation. Self-determination encouraged leaders and liaison people to champion SPIRIT, and tailoring gave them a key selling point – “it’s designed for us”. When self-determination was constrained (e.g. in didactic workshops) it tended to frustrate participants. Tailoring was time-consuming and sometimes required new decision-making pathways, so some agencies found it burdensome and did not make full use of the flexibility on offer (even though they were adamant it was necessary). Where liaison people rejected the premise of SPIRIT (M1) or believed that it was not flexible enough, they put less effort into tailoring and promotion. Leaders wanted choice about when SPIRIT started so it could be used more strategically to complement (or avoid) other activities.

Agency comparisons

Agencies 1, 3, 5 and 6 felt they had enough scope to shape intervention content and did so moderately to extensively. Where there was scope, participants in all agencies contributed to workshop content, increasing its relevance and applicability. Greater involvement seemed to increase receptivity and investment in outcomes. Less effort was made to tailor the intervention in A2 and A4. This may have increased expectations of (and actual) incompatibility. Several liaison people were reluctant to take on their role, and many participants in A2 felt obliged to participate. Despite considerable encouragement, only A5 co-presented a workshop. Two agencies modified non-flexible aspects of SPIRIT: A3 insisted on different participant eligibility criteria to reduce the burden on staff, and A6 requested a hiatus in the intervention while they managed a restructure.
Mechanism 3. The value proposition

**Proposition:** Busy policymakers decide whether to participate based on the intervention’s value proposition. This is determined by what is on offer, at what cost, and how it is communicated. Value can be identified in individual components even where aspect of the overall intervention or trial are rejected. Value is anticipated where the content promises to be useful, stimulating, aligned with local goals, and where there are clear answers about what, why, who and when. Agency-attuned communication is essential.

**Context**

The ideal of research-informed policy was espoused in the wider environment, but each agency had a distinct organisational culture that interpreted this differently. Busy policy staff were juggling competing demands and needed a good reason to take part in non-essential activities. They calculated trade-offs: ‘what can I afford to lose or postpone to make way for SPIRIT?’ All suffered generalised information overload, but many complained about the lack of useful research in their area. The challenge of explaining SPIRIT was exacerbated by: 1. a complex and unfamiliar study design; 2. flexibility (it was being tailored and in flux); and 3. two levels of outcomes: those of the trial (fixed) and those identified by the agency (targeting local goals).

**How did mechanism 3 function?**

A “value proposition” (promised advantage) is a convincing argument about the worth of a strategy that is assessed by prospective users on the basis of perceived costs and benefits. [560] Participants’ view of SPIRIT’s value proposition related to: 1. *Utility* - the content promised to be relevant and applicable, addressing current or future needs. Knowing SPIRIT was locally tailored increased expectations of utility. 2. *Stimulation* - content promised to be interesting. Presenters with “big names”, expert roles, and very senior policy experience piqued interest. 3. *Persuasive marketing* – clear communication using agency-attuned language that emphasised the value of SPIRIT, framed it in relation to agency values and goals, and was disseminated through locally appropriate channels. 4. *Forecasting* – the perceived quality of each intervention activity was used as an indicator of the likely quality of further activities, but only where participants were aware that they were all part of SPIRIT. The value proposition differs from M1 in that it was assessed in relation to each activity - the premise of SPIRIT might be rejected but individual workshops could still promise value.

**How did mechanism 3 generate process effects?**

Where managers saw the value proposition they espoused SPIRIT and encouraged participation, appearing genuine in their efforts. Liaison people supported SPIRIT based on (a) the extent to which it seemed likely to benefit their agency, and (b) whether acting as the LP would benefit or disadvantage them personally. Where liaison people saw the value proposition they went the extra mile to ensure the agency benefitted. Where staff saw the value proposition, participation was moved higher in their list of priorities. Where managers saw the value proposition, participation was moved higher in their list of priorities. The potential benefits of SPIRIT counteracted the burden of data collection (“survey fatigue”) for agencies that had several data collection points prior to the intervention. Initially, SPIRIT marketing was suboptimal: dense, confusing, with poorly attuned “researchy” language (e.g. jargon and acronyms) – policymakers couldn’t see the value proposition. Strategic advice and input from liaison people improved communications substantially.

**Agency comparisons**

Most interviewees in A1, A5 and A6 saw potential value in SPIRIT and so were receptive and inclined to participate. In A3, persuasive internal marketing increased the value proposition from a lower base. Some managers in all agencies encouraged their staff to participate based on potential value. A2 and A4 generally saw less potential value in SPIRIT, and their liaison people put less effort into shaping and promoting it. Some staff in all sites were confused about SPIRIT’s purpose and form, did not know what was expected of them, and entangled the intervention with the trial. A particularly poorly received introductory session in A2 appeared to have lasting effects on perceptions of SPIRIT as a whole, despite some very well received workshops that followed.
Mechanism 4. “Getting good stuff”

**Proposition:** Where policymakers want increased research knowledge, skills or resources, and an intervention offers useful, credible, tangible and stimulating content that facilitates participation, participants are likely to feel that they are getting good stuff. This encourages continued participation, internal promotion, and identification of and intention to use ideas/resources. It also mitigates any data collection demands. Some participants can get good stuff irrespective of their view of the intervention/trial’s premise or broader value, but responses will be shaped by perceptions of current need.

### Context

Policy work tackles complex problems in complex systems. Views about the value of research are contested. Some policy staff were not interested in new skills or ideas, but most wanted and actively sought them. Many were unaware of existing resources in their agencies (human and technical). Each agency had distinct learning norms. Time pressures and reactive practices limited opportunities for policymakers to engage with new ideas and reflect on their practice. SPIRIT demanded considerable amounts of data collection (mostly because of the trial).

### How did mechanism 4 function?

Participants felt they got good stuff from SPIRIT in relation to:
1. **Utility** - content was relevant, applicable, pitched at the right level, partly because presenters had “done their homework”. 2. **Credibility** - presenters with content and practice expertise “got it”, i.e. they understood the constraints and opportunities of policymaking and the need for pragmatism. 3. **Tangibility** - targeted case examples and problem-solving activities made concepts concrete. 4. **Stimulation** - dynamic presenters captured interest and imagination e.g. via compelling behind-the-scenes anecdotes. 5. **Linkage** - interactive activities connected participants to external experts and existing internal resources, and forged intra-agency connections by alerting them to colleagues with expertise or shared work agendas. 6. **Learning congruence** - activities leveraged preferred learning styles. 7. **Reflective space** - workshops provided opportunities for critical thinking. 8. **Orientation** - new staff found that workshop discussions provided insights into how their colleagues view, access and use research.

### How did mechanism 4 generate process effects?

The perceived ‘return on investment’ of participation encouraged (or discouraged) continued engagement in the intervention and data collection, positive word-of-mouth, and receptivity to SPIRIT’s ideas and resources. Feedback form data indicated very positive views, but several interviewees found some content irrelevant, inapplicable or boring and were less inclined to participate in other activities, including data collection. Some spoke negatively to colleagues about their experience, possibly influencing their decision to participate.

### Agency comparisons

Very high numbers of participants across all agencies got good stuff from the majority of workshops. Feedback form results in all sites were extremely positive. Most interviewees engaged with some content, including those who did not anticipate value, and many identified “specific realisable benefits”. Some interviewees believed they would put ideas into practice, and a few had done so, especially newer staff for whom SPIRIT was “formative”. Interview data suggests fewer participants in A2, A3 and A4 got good stuff, but there are only minor differences in their feedback form data compared to other agencies (Additional file 2).

Mechanism 4 focuses on intervention workshops. Perceptions of other intervention components were harder to access. Many interviewees were vague about whether they had seen CEO emails espousing SPIRIT (M8), or if they were receiving weekly updates about resources in the online portal. Those who had accessed the portal said they found it helpful (albeit cumbersome to access due to the need for a password), but were not able to identify any specific use. We did not manage to interview those involved in trialling the commissioned research services in A3 and A4. In the other agencies, the response was mixed—only A5 and A6 were entirely happy about the final product and could identify ways that it would be used. Dissatisfaction mainly appeared to be an artefact of the trial: agencies found it difficult to identify which service and what topic would best meet their needs when required to do so within an externally imposed timeframe. Several participating agencies that struggled to select and tailor their commissioned research service had a history of using the service previously with high levels of satisfaction, but those occasions had been agency-initiated and thus needs-driven.
Mechanism 5. Self-efficacy

Proposition: Where policymakers have some research use skills and are open to further development, self-efficacy can be bolstered through pragmatism, affirmation, modelling, strengths-based dialogue and participation. When this occurs, participants are likely to experience the intervention positively and feel confident and motivated about putting ideas into action.

Context

As analytic thinkers, many (but not all) participants were keen to critique their own, their programs’ and their organisations’ research engagement. Levels of confidence varied: audit survey data suggested the majority of staff lacked experience and confidence in accessing, appraising and using research, whereas 2/3 of interviewees said their skills were reasonable (many had research qualifications and/or experience) but, in most cases, could be improved. Some were apprehensive at the excessive standards of research use that SPIRIT might require. The quality of leadership feedback about staff and agency performance varied.

How did mechanism 5 function?

Self-efficacy refers to people’s beliefs about their capability to perform tasks and achieve goals. [561] Those who feel they have the skills to put ideas and resources into action are more likely to adopt them. [12, 13] Self-efficacy was activated by:
1. Pragmatism - presenters advocated realistic “good enough” goals which assured participants they could achieve acceptable practice standards in using research.
2. Affirmation - participants felt they were building on well-established capabilities partly due to strengths-based audit and mid-way feedback, plus sensitive facilitation in workshop activities, and leaders conveying confidence in their staff.
3. Modelling - high profile experts recounted ‘war stories’, successes and hard-won lessons that countered idealism and echoed messy local attempts to solve problems using research.
4. Experiential learning - trying out tools and systems increased understanding and confidence.
5. Demonstrating expertise - interactive activities enabled participants to contribute valuable local knowledge and skills to SPIRIT content.

How did mechanism 5 generate process effects?

This was one of the least tangible mechanisms but it appeared to have substantial effects on the extent to which targeted policymakers participated in, and were receptive to, the intervention. Many participants who felt they or their agency could use research better were encouraged (and sometimes pleasantly surprised) by SPIRIT content, “We can do this!”. They tended to express enthusiasm for using the ideas and resources they encountered. Some participants who felt overwhelmed by the technicalities or high standards of using research indicated they would probably “leave it to someone else”. Those with greater existing confidence appeared to contribute more actively to the content of SPIRIT activities, thus potentially reinforcing their ‘capability status’ in their own eyes and within the organisation.

Agency comparisons

More interviewees in A1, A5 and A6 than in the other agencies said they were encouraged by SPIRIT. In all agencies, the audit feedback largely supported managers’ understanding of, and confidence in, their staff’s capacity. A minority of participants with less confidence felt overwhelmed (e.g. a very highly rated evaluation workshop in A5 caused a few participants to see evaluation as outside their ability), whereas some others already had high self-efficacy (in all agencies, but possibly A3 in particular). Mid-intervention feedback that showed progress, and which may have supported organisational-level self-efficacy, was not always disseminated within the agency. A few interviewees in A1, A3 and A5 felt that agency-level participation in SPIRIT indicated that their CEO lacked faith in staff capabilities.
Mechanism 6. Respect

**Proposition:** Where participants feel that SPIRIT values them and their work, they are more likely to participate and to be receptive to intervention content. Inferred disrespect or criticism may cause resentment, disengagement and possible generation or reinforcement of negative views about researchers. If the intervention aims to (or is thought to) overhaul valued practices, participants will reject the intervention’s ideas. Closely connected with M1 and M2.

**Context**
SPIRIT participants were experienced professionals with high level skills in information synthesis and analytical thinking. They were specialists in using knowledge strategically in policy processes. Each agency had some history of working with researchers, albeit in quite different ways. In many cases these relationships were regarded as productive, yet stereotypes of arrogant ‘ivory tower’ researchers were pervasive. Some staff had experience of researchers patronising them or treating policy work as unsophisticated. Some agencies had recent experiences of data collection with no feedback.

**How did mechanism 6 function?**
Targeted policymakers felt respected (or, at least, did not appear to feel disrespected) where they regarded SPIRIT as: 1. **Contextually sensitive** - attuned to policymakers’ practices and environment (see M1). 2. **Strengths-orientated** – building on participants’ knowledge and skills. 3. **Supporting practice craft** - SPIRIT accommodated the local “art of policymaking” rather than attempting to impose work practices. 4. **Eliciting meaningful contributions** - participants’ expertise was valued and used in shaping content (M2). 5. **Responsive** - SPIRIT staff and presenters listened, adapted and responded constructively to participants’ queries, concerns and ideas. 6. **Providing feedback** - regular meaningful feedback was given. 7. **Respectful language** - communication and measures were succinct, non-patronising and strengths-orientated.

**How did mechanism 6 generate process effects?**
Respect probably impacted all process effects, but especially no. 3: **Targeted policymakers participate in, and are receptive to, intervention activities.** Most participants appeared to feel understood and heard by the SPIRIT team, which supported engagement, willing participation and openness to what the intervention offered. The few interviewees who experienced aspects of SPIRIT as disrespectful were more critical of the intervention, and expressed more negative views about “typical researchers” and the challenges of researcher-policymaker relationships (countering a key intervention goal of enhancing these relationships). It seemed that these views were usually reinforced, rather than triggered, by their experience of SPIRIT.

**Agency comparisons**
The link between mutual positive regard and process effects was particularly noticeable in A1 where the liaison person responded to concerns on behalf of the SPIRIT team, and interviewees explained they felt more invested in SPIRIT as a consequence. Only a minority of participants indicated they felt disrespected, but we observed occurrences in all agencies except A5. It was due to: perceptions of a patronising or ill-informed premise, perceived assumptions in the measures or in how SPIRIT was introduced, lack of feedback and, on a few occasions, because of a presenter’s comments or delivery style. Belief that the SPIRIT team had made insufficient efforts to understand their needs, practices or context seemed more pronounced in policymakers who did not know the SPIRIT team.
Mechanism 7. Confidence

**Proposition:** Depending on their concerns—which may be affected by data fatigue, previous initiatives, professional interest in the intervention design, and scepticism about researchers—policymakers want to know that a trial in which they participate is scientifically and ethically trustworthy, and that participation is meaningful and poses no threat. Lack of confidence can lead to poor internal facilitation and damage relationships with researchers. Trial activities will affect perceptions of the intervention.

**Context**

The trial required a lot of data collection: six measurement points using three measures, and a process evaluation. Local goal-setting was informed by audit findings from two of the measures. All agencies had previously endured disruptive change initiatives which, in some cases, had little perceived benefit. Many participants had considerable expertise in intervention research and evaluation. Some felt researchers were naive about policymaking and there was a degree of scepticism about what a researcher-initiated intervention could offer.

**How did mechanism 7 function?**

Participants had confidence in SPIRIT when they regarded the intervention, and the trial, as: 1. Valid - (a) the audit and feedback data seemed robust thus (b) the goal-setting was well founded, and (c) subsequent data collection promised to track meaningful change and provide useful findings. 2. Trustworthy - the SPIRIT team were seen to act transparently and non-judgementally, in good faith. Where there were positive pre-existing relationships between agency and SPIRIT staff it facilitated trust in the initial stages. 3. Transparent - it was clear what demands would be made on participants and how data would be used, including when and how outcomes would be communicated. 4. Safe - it was clear that reporting the trial would not compromise individuals or agencies. 5. Effective – the intervention strategies could generate meaningful change in their setting.

**How did mechanism 7 generate process effects?**

Confidence affected the extent to which targeted policymakers wanted to take part in intervention activities and data collection. Liaison people reported that the audit feedback increased leaders’ confidence in, and enthusiasm for, SPIRIT which they communicated to their staff. Participants who lacked confidence in the measures or intervention design, and the small minority who questioned the study’s integrity or safety, expressed discomfort about SPIRIT. Some were avoidant and spoke poorly of researchers. Negative views of the trial were often entangled with the intervention. It seemed that scepticism about researcher-developed content seldom prevented participation, and follow up interview data suggested that attendance at a workshop tended to increase confidence.

**Agency comparisons**

45 leaders took part in audit feedback, and 38 completed evaluation forms. 37 of these (the exception was in A3) answered ‘Yes’ to the following statements: 1. The forum provided clear and accessible information. 2. It provided useful feedback on how we currently use research. 3. The presenter had appropriate knowledge and skills. 4. It gave me confidence that SPIRIT will be tailored to suit this agency. 5. I will encourage my staff to participate in SPIRIT. Despite this high level of confidence, information about the audit was often not disseminated effectively throughout agencies, thus many interviewees did not understand how it had been interpreted or how subsequent data would be used. In A1, a small minority of interviewees initially felt SPIRIT might threaten them or their agency with exposure, but some in all agencies expressed discomfort in “being researched”. Some in A2 questioned the integrity of the study, suggesting it was a business endeavour disguised as research. Liaison people and leaders were instrumental in addressing concerns and increasing confidence, but in A4 and A6 the liaison people themselves questioned the sensitivity of the measurement instruments.
Mechanism 8. Persuasive leadership

Proposition: Leaders persuade policymakers they believe in SPIRIT where the leaders: have credibility as research advocates; support intervention goals visibly, consistently and authentically; articulate the intervention’s value; and model engaged participation. Messages about the need for change must be balanced with assurance of existing capabilities. Expert presenters can be persuasive and inspiring leaders who model values in the wider system.

Context

Participating agencies were bureaucracies with strict hierarchies, but very different infrastructures and numbers of staff. Policymakers in the six agencies had varying levels of respect for their leaders, including different views about the extent to which their leaders valued using research. The professional behaviours of agency leaders in relation to SPIRIT were interpreted locally.

How did mechanism 8 function?

Leaders (managers, opinion leaders and liaison people) were persuasive in support of SPIRIT when: 1. they engaged in value messaging, i.e. they articulated the benefits of SPIRIT including their agency’s scope for and need for increased capacity in using research (M1) in a manner that did not demean current capacity (M5). 2. Respected managers modelled engagement with the interventions’ ideas and resources, thereby connecting SPIRIT to organisational values and priorities. When impressive expert presenters modelled their commitment to research-informed policy and provided examples of its benefits they positioned SPIRIT’s values in the wider policy environment. 3. Agency leadership support was: visible; credible (these managers were known to support and engage in research-informed policymaking themselves); consistent (leaders across the agency conveyed support); and authentic (support seemed genuine). In-person advocacy and participation in workshops was experienced as more authentic than email espousal - actions speak louder than words.

How did mechanism 8 generate process effects?

Persuasive leadership probably impacted all the other mechanisms and process effects, and was a process effect in its own right (see section on mechanism interactions and feedback below). For example, it connected SPIRIT to organisational priorities (M1), increased perceptions of potential value (M3) and trustworthiness (M7), gave staff a positive message about the relationship between SPIRIT and current capabilities (M5), and provided liaison people with a mandate for action (M9). All these impacts, including expert presenters’ advocacy for pragmatic research-informed policy, increased receptivity to the intervention.

Agency comparisons

Leadership support was strongest in A1 and A5. In other agencies managers either lacked credibility as research advocates or their espousal appeared to be perfunctory or inconsistent (e.g. the CEO supported SPIRIT but some managers voiced dissent). Some presenters had insufficient policy-savvy credibility to function as leaders – in all agencies to a small extent, but it was ‘felt’ more in A2 and A4 where resistance was higher. In both A3 and A5 managers’ attendance in workshops bolstered perceptions of SPIRIT’s importance but also inhibited frank discussion by more junior staff. This may have occurred to some extent in other agencies. Forums targeting agency leaders were especially well attended when scheduled within or in lieu of formal management meetings.
**Mechanism 9. Strategic insider facilitation**

**Proposition:** Liaison people are internal staff appointed to coordinate SPIRIT. Where they actively use strategic insider knowledge to support access to and uptake of the intervention via translation, mediation and locally appropriate promotional strategies, they hugely increase awareness and understanding of the intervention, encouraging greater receptivity and participation. Managerial support is required and, even more than other staff, liaison people must anticipate potential value from SPIRIT.

**Context**
SPIRIT required that each CEO nominate a local ‘liaison person’ to coordinate SPIRIT in their agency. Liaison people and managers had their own views about using research in policy work, and about the value of SPIRIT's goals and strategies. In most cases they did not have a say in the agency-level decision to participate in SPIRIT. Beliefs about divisions between the ‘worlds’ of research and policy had currency.

**How did mechanism 9 function?**
In order to be strategic and facilitative, the internal coordination of SPIRIT depended on: 1. **Translation** - liaison people and managers used their ‘insider’ expertise to explain SPIRIT in terms that made most sense to targeted policymakers, illustrating how the intervention intersected with and complemented organisational goals and activities. 2. **Mediation** - liaison people actively identified concerns and worked with the SPIRIT team to resolve them. 3. **Persuasive marketing** - liaison people used local communication channels and creative strategies to “sell” SPIRIT (see M3). 4. **Negotiation** - liaison people and, in one agency, managers, advocated forcefully for adaptations to the intervention that would better suit their needs. 5. **Support** - liaison people were supported internally by managers and externally by a responsive SPIRIT team who provided materials, information and feedback (M6).

**How did mechanism 9 generate process effects?**
Like M8 above, strategic insider facilitation strongly affected other mechanisms as well as SPIRIT’s process effects. High quality facilitation fostered consultative tailoring, continuous engagement and informed participation. Information reached targeted staff in a form that was persuasive and accessible. Problems were identified and resolved so there was greater confidence and receptivity. Poor facilitation led to suboptimal tailoring, confusion about the intervention’s purpose and form, and a poorer local value proposition (M3), resulting in lower levels of participation or unwilling attendance.

**Agency comparisons**
Liaison people in A1, A3, A5 and A6 used creative strategies to tailor and champion SPIRIT, disseminating information through formal and informal channels, e.g. by nominating colleagues to give updates at their team meetings and promoting forthcoming events over coffee. There were transitional lags during changes of liaison person in A3, A4 and A6. The appointment of an enthusiastic and well-supported liaison person in A3 increased survey response rates. Most liaison people spoke highly of the SPIRIT team’s support, but A6 experienced too many points of contact and A4 said there was insufficient guidance. As noted, the liaison people in A2 and A4 did not facilitate SPIRIT to the best of their abilities because they did not believe it was worthwhile (M1, M3, M7). This probably contributed to a lower value proposition (M3) for targeted participants in those agencies, and less confidence (M7). The perceptions, behaviour and impact of liaison people is covered in more detail elsewhere. [329]
Mechanism interactions and feedback

As others have noted, separating interactive processes into discrete mechanisms, while useful for theory-building, fails to reflect their interdependence. [550] Many of the nine mechanisms include related concepts which in some cases may be nested. For example, self-determination (M2) is linked with respect (M6) and may function as a mechanism within self-efficacy (M5).

Figure 9.2 illustrates feedback within our model. This accords with the realist evaluation view that contexts, mechanisms and outcomes are not fixed entities but are contingent on the focus of the current evaluation, i.e. they function as a context, mechanism or outcome in a particular part of the analysis. Thus many of our process effects feed back into and overlap functionally with the identified mechanisms, and may well function as mechanisms when this data is combined with the study outcomes. This is especially pertinent in a process evaluation given that persuasive leadership is a mechanism, despite one of the process effects being Leaders support SPIRIT. This is because we found persuasive leadership to be crucial in activating other mechanisms (e.g. in asserting SPIRIT’s value proposition) and thus in achieving many of the other process effects.

We also concluded that mechanisms functioned on a continuum that encompassed negative and positive expressions. Mechanisms were activated to different extents in each agency and, on occasion, were activated negatively. For example, several interviewees made it clear that mechanisms such as Self-determination, Getting good stuff and Respect were activated negatively when they were instructed by their manager to attend a 2-hour workshop that had no relevance to their work.

Revised program theory

These results enabled us to revise our program theory to reflect contextual contingency which also increases the operational transferability to other interventions and settings (Table 9.3).

Table 9.3. Initial and revised program theory

<table>
<thead>
<tr>
<th>Initial program theory (a-contextual)</th>
<th>Revised program theory (contextually contingent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPIRIT will engage and motivate agency leaders to ‘own’ the intervention using audit feedback, deliberative goal-setting and program tailoring. This agency-driven approach will generate a priority-focused program that offers locally relevant practice support and accommodates differences in agencies’ values, goals, resources and remits. The program will comprise a suite of andragogical activities, tools, and connection across the research-policy divide that provide resources and build knowledge, skills and relationships. It will be supported via modelling and opinion leadership by agency leaders and dynamic external experts. CEOs will promote SPIRIT in their agencies and liaison people will facilitate the tailoring and implementation. These strategies will act synergistically to stimulate and resource participants at different organisational levels, leading to changes in values, practice behaviours and agency processes. This will facilitate increased use of research in policy processes.</td>
<td>Where agencies have an existing orientation to use academic research and are on a trajectory of improved use with perceived room for improvement, SPIRIT will be used to complement or trigger organisational initiatives. Where liaison people and agency leaders believe in the value of the intervention and have confidence in the measures, they will play a pivotal role in tailoring the intervention and championing its goals. Leaders will be motivated by deliberative audit feedback and goal-setting. In all sites, ownership will be increased by greater consultation, collaboration and choice. Agency-attuned communications will be vital in explaining goals, conveying value and addressing concerns. Andragogical activities, tools and connection across the research-policy divide will be valued in all agencies where they leverage existing strengths and address local concerns pragmatically. Staff will make use of these opportunities where they see concrete benefits. Newer staff may benefit most.</td>
</tr>
</tbody>
</table>
Discussion

From participants’ perspective, the most positive attributes of the intervention were: useful (i.e. relevant and applicable) content; high profile experts who delivered pragmatic content and demonstrably “got it”, active participation in intervention activities, and intervention flexibility supported by deliberative audit and feedback that informed goal-setting and customisation. Much of SPIRIT’s implementation fidelity was sound—all the components of the intervention were delivered—but activities were not always as interactive or as participant-driven as intended. Authentic in-person leadership support and committed liaison people were vital mediators, while obstacles included confusion about the purpose of participation in SPIRIT, perceptions of poor alignment with agency practices or priorities, and feeling misunderstood or judged. Previous organisational change initiatives and archetypal views of researcher-policymaker relations sometimes appeared to underpin expectations and frame some of the concerns. The data collection demanded by the stepped wedge evaluation was onerous, and aspects of the trial were often entangled with participants’ perceptions of the intervention. Like many others, we found that pre-existing positive relationships between the agency and those involved in designing and implementing the intervention had considerable facilitative effects. [129, 193, 237] In our case, they helped to activate mechanisms such as respect and confidence.

Implications for intervention improvement

Given their pivotal importance, greater upfront engagement with each agency’s leadership and the nominated liaison person would have been beneficial. Local tailoring and shared decision-making was essential, but challenging for both the agency and the intervention team. For example, it was often difficult for agencies to make strategic use of processes that they had not initiated such as trialling the services for commissioning research. Advice from agencies about how tailoring could be best supported in their context may have been beneficial, but the process of tailoring will always demand time and effort. This reflects the underpinning need for agency leaders to be committed to participation from the start.

Despite being selected for broad similarities, the six participating agencies had markedly different remits, practices and conceptualisations of evidence. SPIRIT’s audit and feedback process was effective in developing a shared understanding of each agencies’ current and desired research use capabilities, but better understanding of their practice goals and values, and greater collaboration in designing the intervention and data collection instruments (which every agency desired) could have sharpened the meeting of minds about what was needed and how to address it. Further, it may have reduced managerial concerns about the demands of the trial and their decision to limit participation in some agencies, which impacted the intervention’s reach. Understanding what participants think about intervention goals, and using their ideas about what should be done to achieve those goals, is usually critical for success. [45]

As noted previously, the realist distinction between intervention activities and mechanisms is crucial for theory-driven evaluation, but it is equally crucial in the development of context-sensitive intervention design and implementation planning. An intervention cannot simply do respect, or deliver self-efficacy, it cannot control the perceived attractiveness of its premise, or make internal facilitators act strategically. Activating these mechanisms is an evolving work-in-progress shaped by personalities, relationships and complex shifting environmental opportunities and constraints. Greater understanding of the mechanisms that generate desired (and undesired) process effects provides helpful guidance, but putting this learning into practice takes creativity, humility and reflexivity.

Our contribution

These findings add to the existing knowledge by surfacing evidence about how policymakers perceived and engaged with different aspects of an intervention trial designed to increase the extent to which they use research in their work. Our realist process evaluation approach goes beyond questions of implementation fidelity and ‘what works?’ to provide a more nuanced and theoretically informed account of how the intervention produced process effects, and why there was such variation across the six policy agencies.

As per Figure 9.2, we anticipate that the intervention’s process effects, and the mechanisms that underpin them, will mediate the study outcomes, but we caution against assumptions that this is a linear predictive
relationship. As realist evaluation adherents point out, there are usually multiple causal pathways in real world interventions, and the best we can do is identify common pathways for particular groups of individuals in particular circumstances; so we concur with McMullen and colleagues that, “there is not, nor can there ever be, a universal implementation model for complex interventions. Site-specific characteristics and realities need to be considered”. [562] However, this consideration need not start from scratch with each new intervention: we can develop an increasingly sophisticated understanding of the conditions that make these outcomes more likely in a given setting. As Pawson argues, “evaluation science assumes that there will be some pattern to success and failure across interventions, and that we can build a model to explain it”. [15] We hope to have made a start in identifying these patterns in a form that will enable others to extrapolate and apply lessons to other interventions and contexts. [15]

**Strengths and limitations of this process evaluation**

Using a realist approach enabled us to: identify and test hypothesised causal mechanisms, evaluate the extent to which SPIRIT activated them, use this analysis to refine the program theory, and identify areas of strength and potential improvement in the intervention and trial design. The identification of underlying causal mechanisms and the development of propositions enhances the utility and transferability of the findings [23, 563] and strengthens the general knowledge base by building on existing theories. The thematic overview of the process evaluation data in Additional file 1, and the inclusion of informing theory in Additional file 2 provide ‘analytical trails’ that support the findings.

Triangulating different types of data obliged us to consider diverse points of view and increased the trustworthiness of our findings. As Wells and colleagues [25] note, “... evaluations need to incorporate multiple methods, multiple sources and multiple perspectives if they are to reflect the context of practice adequately”. We achieved this thanks to the unusually generous appointment of a dedicated process evaluation researcher throughout the study, and the length of the intervention (12 months) and its staggered delivery which gave us considerable time in each agency to test hypotheses at different points in the intervention across six sites. However, we acknowledge this was an exploratory first step and the ideas are yet to be tested by others and in different settings; therefore, at this stage, our findings are only a rough indication of major causal patterns within SPIRIT’s engagement and participation. Further testing and refinement is required.

A limitation was our inability to determine the full range of views and experiences of targeted staff in each agency. Interviewees were sampled purposively for maximum variation of relevant views and experiences, but many declined interviews and it was not always possible to identify substitutes. Others have found similar problems. [233] Consequently, we reached a smaller range of participants than envisaged and so may have missed important views. For example, all the process evaluation interviewees in A4 (11 people with a total of 15 interviews over the duration of the intervention) were either lukewarm or dismissive of SPIRIT, but during outcome measures interviews some A4 participants stated that they welcomed the intervention, and following the trial their CEO said SPIRIT had impacted his agency positively. In all agencies we saw some non-agreement between the highly positive feedback form data and the more critical responses in the interview data. This may be the result of different foci—interviews ranged across the whole of SPIRIT (including its premise, communication and data collection) while feedback forms were workshop-specific—but other factors could be skewed sampling, leading interview questions or the bluntness of the feedback form. The response rate for feedback forms was good with nearly ¾ of attendees (74%) completing them, but it is unclear whether those who did not complete forms differed from those who did and thus what views we might have missed. The direction of this quantitative data was consistent with patterns in the qualitative data regarding a more positive response from agencies 1, 5 and 6, but feedback form responses across agencies and items were so similar that it is likely that the tool discriminated poorly. We used Yes/No statements to maximise response rates from participants who might be rushing to leave, but this was probably too limiting. Certainly, there were many occasions where the free text fields conveyed ambivalence or, at least, scope for improvement, when the scored statements suggested 100% satisfaction. We would use a more sensitive instrument in future.
Reflections on conducting a realist process evaluation

Conducting a realist process evaluation was immensely valuable, but time consuming and challenging. Like others [e.g. 544, 564], we struggled to disentangle aspects of the causal pathways; specifically, to delineate mechanisms from intervention strategies, contexts and outcomes. Realist analysis does not have a step-by-step guide, and it presents a unique tension between ontology and epistemology, so we sometimes struggled to reconcile our search for factual pre-existing mechanisms with the need to take an “imaginative leap” and postulate those mechanisms [565] Three strategies helped: first, scanning appropriate literature and drawing on established theories. For example, the concept of relative advantage [12, 518, 547] was critical for understanding variation in perceptions of SPIRIT and how this linked to the communication strategy. Second, the realist emphasis on counterfactual thinking [44] was very helpful in weighing up the plausibility of different theories. Third, reminding ourselves that causality does not function as discrete components or configurations and that our analysis was intentionally abstracting for the purposes of theory-building rather than attempting to depict reality in all its messy, interdependent glory [see also 550].

Conclusion

This realist process evaluation describes how participants experienced different aspects of a multi-component research utilisation intervention in policy organisations, and why there was such variation across the six implementation sites. We identify nine mechanisms that appeared to facilitate engagement with and participation in the intervention in these settings: 1. Accepting the premise (agreeing with the study’s assumptions), 2. Self-determination (participative choice), 3. The Value Proposition (seeing potential gain), 4. “Getting good stuff” (identifying useful ideas, resources or connections), 5. Self-efficacy (believing “we can do this!”), 6. Respect (feeling that SPIRIT understands and values one’s work), 7. Confidence (believing in the study’s integrity and validity), 8. Persuasive Leadership (authentic and compelling managerial advocacy) and 9. Strategic Insider Facilitation (local translation and mediation). This analysis was used to develop tentative propositions and to revise the overarching program theory. Although our findings are nascent and require further testing and refinement, they indicate areas of strength and weaknesses that can guide the development and implementation of similar studies in other settings, increasing their sensitivity to the range of issues that affect the value and compatibility of interventions in policy agencies.
Chapter Ten: Discussion

“... we need studies to better understand the interacting complexities of change processes.”

Van de Ven and Kangyong, 2011 [566:71]

This thesis focuses on understanding research utilisation interventions in policy agencies. I began by exploring the rationale for such interventions and the many challenges they face; then reviewed 22 interventions that attempted to build policymakers’ capacity to access and use research, and posited some possible causal mechanisms. The rest of the thesis examined a specific research utilisation intervention trial known as SPIRIT. I described how we developed the study’s core definitions, and the methods used to articulate and refine the intervention’s essential elements while simultaneously using them as fidelity indicators. I also reported on the process evaluation findings through two lenses. First, an examination of the views and behaviours of SPIRIT’s internal facilitators, the profound impact they had on local engagement, and what could be done to better select and support these key personnel. Second, how participants experienced SPIRIT, and the possible causal pathways through which intervention strategies brought about process effects.

As outlined in the introduction, the purpose of this research was to investigate the functioning of a multifaceted research utilisation intervention in policy agencies, using SPIRIT as an example, in order to inform the design and evaluation of future similar intervention studies. My three objectives were: (a) To understand what happened in SPIRIT, and why; (b) To contribute to methods in this area; and (c) To contribute to the development of theory is this area. I discuss these objectives in turn:

Understanding what happened in SPIRIT, and why

John Maynard Keynes infamously said, “There is nothing a government hates more than to be well-informed, for it makes the process of arriving at decisions much more complicated and difficult”. [138:196] This quote is often used to mock policymakers for their lack of enthusiasm about the products that researchers want them to use. However, my findings counter Keynes’ view. They suggest that policy decision-making is already complicated and difficult, and that most of the policymakers who participated in SPIRIT wish to be well-informed and to use all forms of evidence as instrumentally possible. However, as we saw in Chapters Eight and Nine, their ideas about what forms of evidence were most useful, and the capacities required to use them, and did not always align with what SPIRIT was offering. This was partly because, in some agencies, the process of developing evidence—and who was involved in that process—was a crucial aspect of the evidence’s legitimacy.

The research findings show that, overall, SPIRIT was very successful in delivering highly-rated content, and in providing a suite of relevant and engaging ‘tasters’ that were especially welcome where agencies were already working towards greater use of research (A1, A5 and A6). SPIRIT seemed to function as an accelerant for this change by providing organisational performance data, and facilitating dialogue about local needs and goals. Then, by providing a range of activities that linked to these goals including: tailored research content, skills workshops, practice in commissioning a research service, research resources and opportunities to talk with researchers. SPIRIT also harnessed leadership support for research use – although this was more successful in some agencies than in others. These activities provided models for further research engagement as well as, in most cases, useful content/experiences in their own right.
Organisational culture was the overarching predictor of engagement style with the trial. SPIRIT was interpreted in relation to the agency’s practice norms and conceptualisation of evidence (strongly influenced by their remit and stakeholder preferences), and the perceived need for improved capacity to use research. The views of liaison people tended to reflect their colleagues’ initial response to SPIRIT, but liaison people’s behaviours then shaped further engagement, affecting what was delivered, how it was promoted, and the extent to which concerns were addressed and opportunities leveraged.

The intervention’s flexibility enabled agencies to tailor content to address needs that were identified both through the audit feedback and via internal processes but, as demonstrated in Chapters Eight and Nine, two agencies experienced SPIRIT as too rigid in terms of its core assumptions and structure to accommodate their preferences. Even where agencies were satisfied with the level of intervention adaptability and the specific activities that they had tailored, they consistently expressed a desire for greater involvement in designing SPIRIT overall.

These findings suggest some overlapping areas for improvement:

- **Strengthened engagement with senior managers and liaison people.** Greater investment in relationship building would serve several purposes. It would increase the research team’s understanding of the agency—its practices, priorities, learning norms and communication style—providing opportunities to fine-tune aspects of the intervention and implementation. It would also increase the agency’s understanding of what SPIRIT could offer and how to get the most out of it, potentially increasing their engagement and activating some mechanisms identified in Chapter Nine (e.g. respect, confidence, persuasive leadership and strategic insider facilitation) which were sometimes lacking.

- **Greater emphasis on agency-driven needs analysis,** including identification of capacity areas that policy staff see as both lacking and necessary. This would likely require that the measures target aspects of capacity that the agencies feel are most important, and do so using instruments that are designed for that purpose. Thinking back to the four domains of capacity identified in Chapter Four, it seems that agencies which were focused on bottom-up systems change might have benefited most from strategies in the Interaction domain so, for example, partnership project trials might be welcomed in those sites. Whereas agencies that wanted to use clinical research more effectively would have chosen strategies in the domains of Access and Skills improvement. Some form of Systems improvement is likely to have been useful in every agency, albeit taking quite different forms depending on the agency’s current infrastructure and resources, as well as its perceived capacity needs. The deliberative dialogues that facilitated audit feedback and goal-setting so effectively in SPIRIT could be extended for this purpose.

- **Greater flexibility and participant involvement in all aspects of the study.** The findings suggest potential benefits in increasing flexibility across the whole trial. This could include aspects of study design such as the intervention timeframe, who participates, and what sorts of options are available. And, as suggested above, the measures could better analyse agencies’ needs but also provide more tailored feedback about progress towards locally determined goals. However, this is a double-edged sword: greater flexibility increases the burden of participation as well as the potential rewards. In the SPIRIT trial it was often hard to obtain timely and efficient input from the busy participants—even those who were enthusiastic—because they were obliged to prioritise work demands and to operate within somewhat sluggish bureaucratic processes. Nevertheless, the more fit for purpose potential that an intervention promises (which would be increased by the combination of improvements suggested here) the more agencies are likely to value participation and therefore to prioritise the investment of time and effort required to hone it for maximised local fit. Importantly, a partnership approach to study design would have a good chance of activating every one of the mechanisms identified in Chapter Nine; a

---

29 These four domains were: Access, Skills improvement, Systems improvement and Interaction.
considerable benefit. It would also support the development of relationships for other research projects which is a key consideration for those who believe that sustained interaction is crucial for advancing research-informed policymaking.

These suggested improvements are not additions to SPIRIT; rather, they are enhancements to strategies that SPIRIT was already attempting to deliver, simply placing greater emphasis on the change principles regarding engagement, flexibility, tailoring for local priorities and agency ownership. Such modifications would present some challenges for evaluation but, as discussed in Chapter Seven, methods are available for capturing the effects of flexible interventions and they are increasingly used in this field, as demonstrated in Chapter Four.

**Reflections on the intervention’s impact**

Ideally, this thesis would have brought together the analyses reported in previous chapters with the findings from the trial’s quantitative impact measures, but this was not possible. First, these quantitative findings have only just become available and are currently being written up. Briefly, the results for the individual self-reported measures (SEER) show an overall improvement in confidence in using research, but not in perceptions of organisational support for research use. The intervention is estimated to have increased the odds of tactical use of research by five times, but there was no significant difference noted for the odds of instrumental, conceptual or mandated use of research. There was some evidence of an increase in accessing primary research, but no change in generating research via partnerships or commissioning.

The results for the measures of organisational capacity (ORACLE) show that, on average, the intervention led to a statistically significant increase in: the availability of organisational programs to train staff in using research; mechanisms that help strengthen staff relationships with researchers; and the overall score of organisational capacity. There was some evidence of an intervention effect regarding the presence of systems/methods to generate new research to inform the agencies’ work, but no statistically significant increase was found in methods to ensure adequate evaluation of organisational policies and programs, or in processes that encourage or mandate the examination of research in policy and program development.

Second, there was not enough statistical power for comparison between the agencies, including differences between the ‘engaged’ and ‘non-engaged’ agencies, so it is not possible to explore the relationship between the variation identified in this thesis and the study outcomes. This is unfortunate because it means I cannot examine the extent to which process effects aligned with study outcomes (e.g. did agencies where more desirable process effects were observed have better outcome scores?), or explore which process effects may have made the most important contributions to the outcomes. This is undoubtedly a limitation.

But it would have been a mistake to expect a linear alignment. Intuitively, positive change seemed more likely in the agencies that embraced SPIRIT (A1, A5 and A6), and least likely in the agencies where the liaison person rejected the premise of SPIRIT (A2 and A4) or where participants saw little room for improvement (A3). But despite the tendency for participants in A2 and A4 to reject the premise of SPIRIT, many still attended workshops that they rated well on feedback forms, and took part in process evaluation interviews where they talked enthusiastically about the various forms of evidence that inform their work. There was a

30 Guidance in this area has increased since the paper in Chapter Seven was published. See, for example, a recent report by the Overseas Development Institute: [https://www.odi.org/sites/odi.org.uk/files/resource-documents/10361.pdf](https://www.odi.org/sites/odi.org.uk/files/resource-documents/10361.pdf)

31 These findings derive from the following survey item—*It is valuable to use research in policy or program work to:* 1. Identify issues that require a policy or program response, 2. Understand how to think about issues, 3. Decide about content or direction of a policy or program, 4. Persuade others to a point of view or course of action, 5. Design the implementation or evaluation strategy for a policy or program, 6. Monitor implementation or evaluate the impact of a policy or program, 7. Meet organisational requirements to use research. Options 1, 3, 5 and 6 are were coded as instrumental use, option 2 as conceptual, option 4 as tactical, and option 7 as mandated use.
sense that SPIRIT was amplifying the conversation about evidence-informed policymaking in all agencies, irrespective of the extent to which there was agreement about what this term meant, and that this conversation might provide the impetus for agency-led initiatives. Indeed, my research captured some of these unanticipated impacts, two of which are summarised in box 10.1.

**Box 10.1. Two examples of unanticipated (and hard to measure) impacts**

In A2 (an ‘unengaged’ agency) two senior managers argued that SPIRIT would have been more useful if it had supported improved implementation of the agency’s programs via frameworks such as Kotter’s eight steps for leading change. [567]. These managers had discussed this deficit following a SPIRIT workshop and decided to address it themselves. One manager said she had planned a workshop on this topic with her staff, and the other had already run one, “Since then, I ran a session in-house... which just broke it down to basic principles.... How do you look at process implementation? How do you evaluate that?”. Therefore, despite the predominant view in this agency that SPIRIT’s form and assumptions were incompatible with agency practice, the intervention triggered some potentially useful ideas that were being put into practice.

Many A3 staff argued that their agency had little room for improvement in using research. However, following an evaluation workshop and leaders’ forum, several members of the evaluation team and other managers commented that A3 was making suboptimal use of its internal evaluation. In interviews, these policymakers reported being more assertive in project meetings about using evaluation findings, and in subsequent SPIRIT workshops several of them raised concerns with colleagues about their use of evaluation. There was no indication that SPIRIT had increased skills, but it encouraged reflection about the agency’s use of internal knowledge which revealed some self-identified room for improvement, and it provided a forum that brought different organisational teams together for discussion. Within such a research-orientated work culture, this seemed to be sufficient to generate proximal change.

The increase in confidence and the projected increase in tactical use of research (i.e. using research to persuade others) in SPIRIT’s outcomes are both somewhat surprising, and suggest opportunities for further investigation. Increased confidence is unexpected given that training workshops were relatively brief (1½ to 2 hours per topic) and were not accompanied by embedded support structures such as mentoring and practice assessment that we saw working effectively in Chapter Four. However, the much-appreciated emphasis on pragmatism that characterised many workshops may have contributed to a feeling of achievable. Confidence corresponds poorly with actual competence [568], but is an important aspect of self-efficacy and thus of people’s openness to practice change [7, 569], so this may bode well for greater use of research in the longer term. It would be useful to better understand the relationship between confidence, capacity building strategies and engagement with research.

The projected increase in tactical use of research is also unexpected. An intervention in policy agencies might well gain traction by framing itself as providing argumentative tools, but that was not prominent in SPIRIT. However, this finding could be seen as a success. As demonstrated in Chapter Three, argument is the predominant discourse in policymaking, therefore making tactical use of research can be highly productive in terms of utility and impact. [84] As SPIRIT progressed, some participants said they experienced greater organisational expectation to use research in their work. This fits with the increase in overall ORACLE scores (above), and may account in part for attributing more value to research as a persuasive tool within internal processes. Investigation into the scope and process of tactical research utilisation might provide more nuanced insights. This is warranted not least because using research as argument rather than as data or ideas [84] risks selectivity and misrepresentation [33], so we cannot afford to conflate increased use of
research with improved use.\textsuperscript{32} Intervention design, therefore, faces a challenging balancing act between recognising and working with the use of research as an instrument of influence, but not promoting use at the expense of what Jones calls the “right kind of influence”. [88:29]

From these findings, it is uncertain whether or not SPIRIT can generate sustainable change. Hawe and colleagues conceptualise “interventions as events in systems that either leave a lasting footprint or wash out.” [67:270] To leave a lasting footprint an intervention needs to do some of the following: generate new roles and relationships; establish new activity settings; embed actions, ideas or tools in routine practices, organisational systems and workplace conversations; and attract ongoing resources such as time and funding. [67] SPIRIT evidently generated new roles and relationships via the liaison people. The role itself was temporary, but in some cases the change in status and new patterns of interaction may have enduring effects. Some new relationships between policymakers and researchers were established and were being pursued in at least two agencies (A5 and A6). Skills workshops did not establish new activity settings—every organisation already had similar learning forums—but in some agencies the more targeted and conversational ‘research exchanges’ (where policymakers and researchers met to discuss the evidence on a priority topic) seemed to offer a different and productive model. There were indicators that some SPIRIT-related actions, ideas and tools had seeped into routine practices, organisational systems and workplace conversations. Examples include: leaders using new language to talk about research use (A1), mandating the consideration of research in policy and program planning (A1 and A5), changing default search strategies to include recommended resources (A1 and A3), and the integration of commissioned research into program development (A5 and A6). It seems likely that SPIRIT will have some legacy, even if it is not to the extent or in the manner that the research team hoped for.

**Contribution to methods in this area**

This thesis makes a unique contribution to the literature by using an in-depth process evaluation of a novel research utilisation trial in policy agencies to develop new methodological approaches for understanding how and why complex interventions function as they do. As Chapter Four illustrates, there have been few policy-focused utilisation trials, and none has reported on process effects in as much detail as SPIRIT. This work demonstrates how qualitatively-orientated process evaluation can unpack the ‘black box’ of implementation and contribute significantly to capturing its process effects. Our understanding of what happened in SPIRIT would have been poorer without this work. Despite its increasingly vocal advocates [e.g. 5, 6, 17, 25, 400], process evaluation is given little attention in the intervention literature. This is a serious problem, and suggests that the research community continues to underestimate its considerable benefits as a complement to outcome evaluation. [2]

The thesis provides worked examples of methods which support the development of theory-informed process evaluation. They include practical methods for: articulating and assessing fidelity in complicated intervention trials, conducting a context-sensitive mixed methods process evaluation, and wrangling large amounts of diverse data. Where possible, tools and examples are included (e.g. coding frameworks, interview questions, case study structures and examples of evolving analyses). The findings provide evidence of the rich explanatory findings that such methods can deliver. Chapter Nine extends theory-informed process evaluation methodology further through the application of realist approaches. Realist process evaluation is still rare\textsuperscript{33}, so this work adds usefully to a burgeoning field.

\textsuperscript{32} An example of this is Waldman’s (2014) study of policymaking in a British government department which found that staff were highly research-engaged but that symbolic “justificatory” use of research was rife.

\textsuperscript{33} A Google Scholar search for “realist process evaluation” in May 2017 returned 24 results, of which two were protocols (Randell et al. 2014, and Hensel et al. 2016), two were conference abstracts, and only two were published accounts of realist process evaluations (van der Zijpp et al. 2016, and Moore et al. 2016). See also Masterson-Algar and colleagues’ (2014) work in applying realist methods to fidelity assessment.
Contribution to the development of theory in this area

Perhaps the most important (but least tangible) contribution this thesis makes is in furthering our understanding of the dynamic and highly situated connections between policymakers’ information needs and practices, and different kinds of research utilisation intervention and implementation strategies. This has implications for future research utilisation interventions in policy agencies, and maybe even for how we think about policymakers’ diverse context-bound perceptions of research and researchers more generally. This learning is brought together in provisional propositions which allow others to apply lessons from this research in their own initiatives and settings.

Many of the constructs that informed this work (presented in Chapter Five), have been substantially advanced. In fact, following this research, most of the original constructs now seem obvious, over-generalised and unhelpfully abstracted—useful for suggesting domains of inquiry for a process evaluation but offering little guidance for intervention design, and shining very little light on the contextual specifics required for theory-building. Table 10.1. provides an overview of ‘where we are now’ in relation to those constructs which are now more nuanced and have clearer implications for subsequent interventions.

Given that this is an applied field, I have attempted to make this work practical, using somewhat different methods in each study. In Chapter Six, pragmatic methods for developing definitions are suggested. Chapter Seven presents a worked example of theoretical fidelity assessment and advice about applicability in other contexts. Chapter Eight focuses on interactions that were specific to SPIRIT but makes practical recommendations that apply to other interventions. And Chapters Four and Nine use realist approaches to identify tentative middle range theory that maximises the transferability of findings and provides points of departure for others to build on. Where a realist approach is used I have supplied additional files so that (a) those who struggle to make meaning through a realist lens can still access key take home messages, (b) the underpinning theories are more accessible, and (c) so that readers can see more of an ‘evidence trail’ and conduct their own plausibility assessment of the results.

This work draws on theory from political science, organisational change, behavioural psychology, evaluation theory, systems thinking and implementation science to delve deeply into questions about research utilisation in policy agencies, and how interventions in this space might fare. The use of different lenses to explore facets of these phenomena illustrates an overarching message: that theory should be eclectic when investigating, and intervening in, complex social processes such as research utilisation which straddle individuals, organisations, disciplines, relationships and systems. Consequently, the thesis is transdisciplinary in that it (a) takes account of complexity, (b) addresses diverse perceptions of the core issue, and (c) takes a pragmatic approach to produce practically relevant knowledge.

Together, the contributions outlined above offer empirically and theoretically grounded guidance for the design, implementation and evaluation of future intervention studies in this field and beyond.
Table 10.1. Tentative constructs used to guide the process evaluation, and reformulations based on findings

<table>
<thead>
<tr>
<th>Tentative pre-intervention constructs</th>
<th>Where we are now (post-process evaluation reformulation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Organisational culture mediates interventions; agencies where research is valued will make better use of intervention resources</td>
<td>Of course organisational culture mediates interventions, but the issue is not so much whether research is valued per se as what makes (or could make) different forms of research valuable in this context? Perceived alignment between what the agency valued and what the intervention targeted generated more enthusiasm for SPIRIT, and greater self-reported engagement with its ideas and resources. Incorporating each practice culture better in the intervention design and audit might have led to increased perceptions of value, stronger engagement and improved outcomes.</td>
</tr>
<tr>
<td>2. Organisational role affects what research is valued and how it is used</td>
<td>Yes, but other meso level factors (e.g. team culture, practice norms, institutional supports,) and shifting circumstances at both meso and macro levels (e.g. political climate) are just as important. Situational analysis and reflexive implementation and evaluation (aided by insider nous) are required to understand and respond to these conditions.</td>
</tr>
<tr>
<td>3. The credibility of workshop presenters affects how policymakers engage with content</td>
<td>Yes, but as demonstrated in Chapters Four, Eight and Nine, what is credible will vary. This research suggests that hands-on expertise was most important in some sites, while academic credentials impressed participants in the more research-engaged agencies. Agency learning norms and participants’ backgrounds, as well as the almost universal need for relevance and applicability, also appeared to affect how content was assessed.</td>
</tr>
<tr>
<td>4. Involvement in tailoring strategies increases local relevance and applicability, and a sense of ownership</td>
<td>Findings support the view that people prefer to implement and participate in plans that they have made themselves [566]; however, as shown in Chapter Eight, the promise of tailoring can backfire if the intervention does not permit as much flexibility as participants desire. Participants must be prepared to dedicate time and effort to tailoring to reap the benefits, and may require guidance about what can be achieved and through what methods.</td>
</tr>
<tr>
<td>5. Policymakers’ existing values, knowledge and experience will shape if/how they use research and how they engage with a research utilisation intervention</td>
<td>This is one of the most obvious yet least straightforward propositions, mainly due to contextual variation. For example, valuing academic research did seem to increase perceptions of SPIRIT’s value but only where the agency was regarded as amenable to change and had room for improvement, and where current roles and tasks were considered likely to benefit from using research (Chapters Four and Nine). These concepts are familiar from much implementation and research utilisation literature e.g. [11-13, 79, 134, 147, 571]</td>
</tr>
<tr>
<td>6. How participants perceive an intervention will strongly affect how it functions in each site</td>
<td>This is clearly supported by findings reported in Chapters Eight and Nine; however, it is a strangely underestimated factor in much intervention research. [532] Importantly, perceptions were not fixed: in many cases they changed over the course of the trial. Change was largely positive thanks to the research team’s responsibility to local conditions. This reinforces the importance of thinking about an intervention as an active process in an evolving system, rather than as a suite of deliverables.</td>
</tr>
<tr>
<td>7. Interactivity, practice opportunities and other adult learning techniques enhance engagement, learning and self-efficacy</td>
<td>Findings supported the link between interactivity and engagement (and, by implication, learning), but with caveats. First, interactivity was not always achievable—presenters with context expertise were not necessarily skilled facilitators or inclined to follow protocols (Chapter Seven). Second, the content had to promise sufficient reward to warrant input by policy participants (chapters Eight and Nine). And third, it was not universal—while interactivity was strongly appreciated and enormously valuable overall, on a few occasions, tired participants said they would have preferred to ‘receive’ content rather than co-creating it.</td>
</tr>
<tr>
<td>8.</td>
<td>Managers are key drivers of organisational change; their espousal of a research utilisation intervention and modelling of research engagement will be crucial</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>9.</td>
<td>Audit feedback can increase awareness of research capacity needs which may, in turn, motivate managers to support a research utilisation intervention</td>
</tr>
<tr>
<td>10.</td>
<td>Archetypal differences between researchers and policymakers may affect research utilisation interventions (e.g. they could be perceived as patronising or lack local applicability)</td>
</tr>
</tbody>
</table>
Future research directions

We need to understand the processes that we are trying to influence

An important concern in Chapters Two and Four was that we continue to have suboptimal understanding of how research and other forms of evidence are actually used in policymaking. [35] But there are opportunities for addressing this. Two related fields of inquiry seem to offer the richest potential. They are:

1. Unpacking the process of policymaking and 2. Investigating the social life of research within policy processes.

1. Unpacking the process of policymaking focuses on the day-to-day practice of policymakers. Brown and Duguid [76] draw attention to the centrality of process in understanding work practices. They cite Bourdieu’s (1973) differentiation between modus operandi (the way a task looks to someone working on it as it unfolds over time when “many of the options and dilemmas remain unresolved”) as opposed to opus operatum (the way a finished task appears in hindsight). Opus operatum, they argue, focuses on the task and glosses over the process of doing which is structured by changing contextual conditions. This means that as a work process becomes more complex, our post-hoc depiction of it increasingly obscures what actually needs to be done. It seems then, that close attention to modus operandi is needed to understand how research is used (and can be used) in policymaking. Weiss makes a similar point, noting that there is an important difference between policymakers’ espoused criteria for useful research and their in-use criteria. [162]

As we saw in Chapters Nine and Ten, some SPIRIT participants felt that the intervention design and measures failed to recognise key aspects of local modus operandi and in-use criteria. Unfortunately, SPIRIT has not advanced this significantly; research techniques such as those used in the measures and process evaluation can only capture opus operatum and espoused criteria. Most other common approaches to investigating research utilisation have the same limitations. Therefore I share the view of Oliver et al. [35] that ethnography is an important future research direction for understanding knowledge utilisation in policymaking. A similar argument can also be made for understanding how interventions work in complex contexts [265], but note the potential challenges of ethnography in the context of trials described in Chapter Five.

Ethnography has already made some important contributions to this field. [38] For example, Stevens [87] found that a section of UK policymakers he worked with had a normative commitment to research-informed policymaking and wove research argumentatively into the narratives they used to advance policy proposals, but that they shed caveats to give the illusion of greater certainty, failed to question organisational assumptions, avoided conflict with current policy narratives, and developed “totemic” policies that neglected compelling but inconvenient data about social inequity. Greater understanding of these processes provides valuable information about organisational strengths and weaknesses that can inform internal change initiatives and the design of intervention trials. It also echoes the point made above about the danger of assuming that research use per se is a success.

2. Inquiries into the social life of research within policy processes take the reconfiguration and movement of research as the unit of analysis. This recognises that research findings are always adapted rather than adopted [28]; given legitimacy rather than having it [391]; and that the ideas from research, rather than the findings-as-reported, are what travels. [92] The movement of research ideas, and the ‘journey’ metaphor, have been pursued by Smith [92] who demonstrates that how research ideas are conceptualised shapes the journey they will take. Evolutionary biology has been used to describe the survival of research ideas that fit in policy, including the notion that some research ideas may function like memes (patterns of culturally
embedded information that pass ideas from one person to another [572]). [95, 573] Such an approach would demand an ecological, systems-informed view of policymaking. [170, 474]

A promising program of research investigating the social life of research (including the social life of ideas from an intervention) might include network analysis [67, 77, 170]; tracking relationships and impacts [67]; and, as already suggested, ethnography. Gabbay and le May provide a powerful example of using ethnography to reveal the social life of research. Their study of how GPs respond to clinical guidelines found that they integrate information from guidelines through an ongoing process of collective scrutiny, surfacing their peers’ tacit knowledge, and reference to practice experiences, transforming this amalgam of information into ‘mindlines’. [31, 74, 574] Like the work by Stevens [87], above, it shows (a) that the social arena of professional life shapes what forms of information are considered and how they are used, and (b) that knowledge is enacted, i.e. it is formed through an evolving process of contextualised meaning-making. Indeed, Gabbay and le May use the term “knowledge-in-practice-in-context” to describe the interaction between work practices, information utilisation and local circumstances, arguing that professional knowledge exists only through doing. They explain,

“If knowledge-in-practice-in-context is something people do, not something they have (i.e. it is performative, not substantive), then a fundamental change is needed in the prevalent assumptions about the nature of evidence and knowledge for policy and practice. Evidence is not just research-based knowledge that one either has or does not have. Knowledge transfer is not just a matter of handing someone a parcel of new evidence, however easy it is to open and however carefully chosen the contents. Knowledge translation is not just a question of someone from one world with some new evidence getting someone from a different one to understand it. Knowledge utilization is not just about applying or failing to apply evidence when one has found it. They are all complex social processes of knowledge transformation.” [31:199]

Those of us who hope to advance the use of research in policymaking must respond strategically to this growing understanding of research utilisation as a dynamic practice of knowledge transformation.

There is also scope to expand researchers “policy literacy” [575] more generally. Research in this space is likely to focus on evaluating initiatives opportunistically rather than formal trials, and it must be sensitive to contextual considerations highlighted in this thesis such as values, beliefs, relationships and systems dynamics. Promising strategies already in progress include researcher secondments in policy agencies [576], policy literacy courses [577], and ongoing linkage and exchange forums [317, 578] (although, as we saw in Chapter Four, the implementation and impacts of this approach are far from straightforward).

**We need to develop intervention strategies that tackle the processes we are trying to influence**

In the six agencies that participated in SPIRIT, the intervention was judged in a similar way to that in which research was judged: in terms of local and temporal utility. The value of the intervention, then, is not intrinsic but determined by its potential capacity to make a difference within specific circumstances, and it is shaped by interpersonal interpretations within a cultural frame. Thus value is conceived in terms of local interaction, but is also arrived at through local interactions.

This has several implications for research utilisation interventions. First, framing research questions in terms of knowledge transfer [e.g. 311]^{34} risks focusing on products at the expense of processes. Rather, we need to examine how processes shape the movement, meaning and utility of products (and, perhaps more

---

^{34} Lavis et al. (2003) pose the following research questions: “what should be transferred; to whom should research knowledge be transferred; by whom should research knowledge be transferred; how should research knowledge be transferred; and, with what effect should research knowledge be transferred?”
importantly, of the ideas that stem from them), identify interactions that may be amenable to intervention, and test ways of targeting those. Promising endeavours include building the capacity of internal opinion leaders to champion research, and of policy agencies to develop their own syntheses, using trusted insiders and boundary spanners as information conduits, and further development of research-policy partnerships. This is not to suggest that the characteristics of research products are irrelevant, far from it, just that we already know a lot about what tends to be most useful in terms of focus, content and packaging [e.g. 28, 146, 579, 580], and further information would likely emerge from an investigation of processes.

Second, we have seen that workplace cultures are orientated to particular types/blends of evidence and particular ways of using it, and that factors such as practice norms, leadership, professional networks and stakeholder interests contribute to this culture. Critical questions about research utilisation, then, are less about ‘What?’ or ‘How much?’ and more about ‘Who?’, ‘Why?’ and ‘In what ways?’. Correspondingly, research utilisation interventions may need to focus more on understanding local practice rationales and responding to those, rather than assuming that certain forms of research are exemplary evidence and should be used in particular ways. [35]

Lastly, these social and process factors were also enormously important in shaping how participants perceived and engaged with SPIRIT and the other interventions reviewed. This suggests that qualitatively-orientated process evaluation has a vital role to play in answering questions about how and why interventions function as they do in different circumstances. This thesis is evidence of the illuminating information that can be gathered during the implementation of an intervention using these methods.

**Conclusion**

This thesis provides a rich description of activities, views and experiences across the SPIRIT trial. Its real strength, however, is that it is explanatory as well as descriptive, so it not only enhances our knowledge of what happened, but provides empirically and theoretically grounded accounts of why it did so. The contents advance conceptual understandings of how research utilisation interventions are likely to function in policy agencies under different conditions, and provide methodological guidance for evaluating such interventions. I hope this work will contribute to broader efforts to understand and influence the use of research in policymaking. Former US national security adviser Jake Sullivan remarked that, "Public policy is a study in imperfection. It involves imperfect people, with imperfect information, facing deeply imperfect choices, so it’s not surprising that they’re getting imperfect results." [581] Better use of relevant and robust research in policy processes will not transform this scenario, but it may lessen information imperfections and help clarify some policy choices. My thesis is a small contribution to this worthwhile effort.
Thesis references


122. Erren TC, Shaw DM and Grob JV. How to avoid haste and waste in occupational, environmental and public health research. Journal of Epidemiology and Community Health. 2015; 69(9): 823-5.


135. Chapman S. "One hundred and fifty ways the nanny state is good for us." The Conversation (2nd Jul), 2 Jul 2013.


Lanham HJ, Leykum LK, Taylor BS, McCannon CJ, Lindberg C and Lester RT. How complexity science can inform scale.


Galea S. On knowledge and values. Dean’s Note. Boston University, School of Public Health, 2016.


197. Stafford T. “How curiosity can protect the mind from bias.” BBC Future, 8 Sep 2016.


201. Hannes K, Behrens J and Bath-Hestall F. There is no such thing as a one dimensional hierarchy of evidence: a critique and a perspective. Cochrane Colloquium: 23. Vienna, Austria2015.


368. Rothman AJ. “Is there nothing more practical than a good theory?”: Why innovations and advances in health behavior change will arise if interventions are used to test and refine theory. *International Journal of Behavioral Nutrition and Physical Activity*. 2004; 1(1): 11.


Vohnsen NH. Absurdity and the sensible decision: implementation of Danish labour market policy (PhD thesis in anthropology). Aarhus University, 2011.


423. de Leeuw EJ. Health policy: an exploratory inquiry into the development of policy for the new public health in the Netherlands. 1989; Maastricht: University of Maastricht.
430. Feldman PH, Nadash P and Gursen M. Improving communication between researchers and policy makers in long-term care or, researchers are from Mars; policy makers are from Venus. The Gerontologist. 2001; 41(3): 312-21.


Francis JJ, O’Connor D and Curran J. Theories of behaviour change synthesised into a set of theoretical groupings: introducing a thematic series on the theoretical domains framework. *Implementation Science*. 2012; 7: 35-.


497. Grimshaw JM, Eccles MP, Greener J, et al. Is the involvement of opinion leaders in the implementation of research findings a feasible strategy. Implementation Science. 2006; 1(3).


504. QSR International Pty Ltd. NVivo qualitative data analysis software: Version 10 2012.


516. Evans BA, Snooks H, Howson H and Davies M. How hard can it be to include research evidence and evaluation in local health policy implementation? Results from a mixed methods study. *Implementation Science.* 2013; 8.


523. Thomas DR. Feedback from research participants: are member checks useful in qualitative research? *Qualitative Research in Psychology.* 2017; 14(1): 23-41.


529. Thomas DR. Feedback from research participants: are member checks useful in qualitative research? *Qualitative Research in Psychology.* 2017; 14(1): 23-41.


Thorne S. *Interpretive description*. Walnut Creek, CA: Left Coast Press, 2008.


Appendices

Appendix 1: Glossary and acronyms
Appendix 2: Models of research utilisation
Appendix 3: Review characteristics and search strategies
Appendix 4: Overview of included studies
Appendix 5: Context and process
Appendix 6: Early process evaluation interviews (general participants)
Appendix 7: Post-intervention process evaluation interviews (general participants)
Appendix 8: Headings used in the Framework Analysis matrices
Appendix 9: Chapter Six Manuscript
Appendix 10: Chapter Seven Manuscript
Appendix 11: An example of how selected essential elements changed during SPRIT
Appendix 12: Chapter Eight Manuscript
Appendix 13: Chapter Nine Manuscript
Appendix 14: Descriptive overview of results
Appendix 15: Supporting theory for the realist process evaluation analysis
Appendix 16: Publications arising from this thesis
Appendix 17: Acknowledgements
Appendix 1. Glossary and acronyms

Glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Meaning/use in this thesis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constructionism Vs constructivism</strong></td>
<td>These terms are often used interchangeably. The distinction is made in this thesis because I draw heavily on Crotty’s [1] work in Chapter 2. Crotty argues that constructivism is associated with a radical relativist epistemology whereas the focus here is on intersubjectivity and the social development of knowledge for which constructionism is the best term.</td>
</tr>
<tr>
<td><strong>Context</strong></td>
<td>Context is used in two ways in this thesis. Mostly, the term is used loosely to mean “the set of circumstances or unique factors that surround a particular implementation effort”. However, when talking about context within a realist study the term is used more precisely, as realists do, to include three levels: (a) meso-level phenomena such as organisational culture and systems, (b) wider macro-level phenomena such as political climate, health services infrastructure and institutional trajectories of change, and (c) people’s current resources and reasoning, and patterns of interaction. See Appendix 5 for more detail.</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>The products of research or rigorous inquiry, broadly. Given the emphasis on qualitative data in this thesis, data is considered a ‘mass noun’ rather than a ‘count noun’: a mass, like water, traffic, history—or qualitative data—cannot be counted. Hence I say “this data” rather than “these data”.</td>
</tr>
<tr>
<td><strong>Evidence</strong></td>
<td>Evidence for policy is contingent. It often comprises multiple forms of information from diverse sources, and may be formulated to reflect particular interests. What constitutes evidence for one person may be seen as anecdote or ideology by another. Consequently, evidence is used as an idea in this thesis, and a loaded one at that, rather than as a concrete noun like research. See Chapter Three for more detail.</td>
</tr>
<tr>
<td><strong>Fidelity assessment</strong></td>
<td>The purpose of fidelity assessment is to ascertain “the degree to which an intervention or procedure is delivered as intended”. This is achieved by operationalising the intervention theory and monitoring the consistency and congruence with which it is implemented. Often associated with adherence to a standardised intervention, I follow Hawe and others in applying the principles to flexible interventions and theory-informed evaluation.</td>
</tr>
<tr>
<td><strong>Knowledge</strong></td>
<td>There is a wealth of philosophical and practical literature about the nature of knowledge and knowing and its relationship to policy that is outside the scope of this thesis [e.g. 12, 13-16]. Knowledge is used pragmatically here to include research findings and ideas, and other forms of information that might be treated as authoritative within policy processes, e.g. local surveillance data, expert advice and practitioners’ experiential nous. However, I note that others use knowledge more precisely to indicate information that is integrated, so that information becomes knowledge only when people are able to “collectively and/or individually combine it with their own experience, skills, ‘intuition’, ideas, judgements, motivations and interpretations” [Ahmed in 17:102]. This concept influences my thinking (especially in Chapter Ten), but not my terminology.</td>
</tr>
<tr>
<td><strong>Knowledge transfer, translation, exchange, or mobilisation</strong></td>
<td>Akin to the debate about evidence-based policy, the discourse about how knowledge is shared has moved from a linear rational model towards an increasingly multi-directional and complexity-attuned model where knowledge is produced and becomes meaningful through social processes. The evolution of terms from transfer to translation to exchange and now to mobilisation illustrates this well. I use the term knowledge mobilisation in this thesis because it is the most fluid and inclusive option, concerned with “moving knowledge into active service for the broadest possible common good”.</td>
</tr>
<tr>
<td>Mechanism</td>
<td>Realists focus on an intervention’s underlying theory rather than its activities; specifically, on the causal mechanisms that generate outcomes in particular contexts. [4, 20] Mechanisms are what makes an intervention work: “They are not the observable machinery of program activities, but the response that interaction with a program activity or resource triggers (or does not trigger) in the reasoning and behaviour of participants.” [21]</td>
</tr>
<tr>
<td>ORACLe</td>
<td>One of the three outcome measures used to evaluate SPIRIT, along with SEER and SAGE. ORACLe was a structured interview with a senior member of staff from each agency about their organisation’s capacity to use research. Transcripts were scored using an algorithm. See Chapter Five.</td>
</tr>
<tr>
<td>Policy</td>
<td>Policy is used broadly to include any formal statement or action plan developed by a government agency or statutory body in response to an identified problem. This includes state-wide or national legislation, policies, programs, directives, protocols, guidelines, service models and standards (Chapter Six). Policy processes include the myriad activities that surround this work: formal and informal, intentional and unplanned.</td>
</tr>
<tr>
<td>Policy agency</td>
<td>These are government organisations, or government-funded organisations that have statutory powers, i.e. the authority to exercise powers on behalf of the government, including forming regulations and implementing legislation and other policies.</td>
</tr>
<tr>
<td>Policymaker</td>
<td>Policymakers work in policy agencies and are directly involved in the development, implementation or evaluation of policies or programs, or decisions about their funding. Other than in Chapter Four, three important policy groups are excluded from this definition because they have different powers and responsibilities and were not targeted by SPIRIT. They are: Service managers (such as those running area health services), people working for NGOs who often play important roles in policy coalitions, and elected ministerial officials and their staffers. The review in Chapter Four is more inclusive, as explained therein.</td>
</tr>
<tr>
<td>Process effects</td>
<td>These are proximal effects that are likely to influence study outcomes or be of evaluative interest for other reasons (e.g. they impact on relationships between the researcher and policy agencies, or help explain local adaptation processes or unexpected variation in implementation). Desired process effects are those that the investigators consider to be prerequisites for a successful intervention. Chapter Nine uses process effects in lieu of outcomes within a realist Context+Mechanism=Outcome analysis.</td>
</tr>
<tr>
<td>Realism</td>
<td>“Realism is a methodological orientation, or a broad logic of inquiry that is grounded in the philosophy of science and social science... it asserts the existence of a reality which is independent of our senses, but which we can only discover through our senses.” [22] It is a theory-driven approach that strives to explain how and why interactions between intervention strategies and contexts generate observed outcomes. Chapter Two presents my position on realism in more detail.</td>
</tr>
<tr>
<td>Realist</td>
<td>Realists here means those who follow the scientific realist position advocated by Pawson and Tilley and others in the RAMESES group. I do not include those in the critical realist school, nor do I use the term realist in a ‘lay’ manner to suggest people who refuse to view the world through rose-tinted glasses.</td>
</tr>
<tr>
<td>Research</td>
<td>In this thesis research incorporates evaluation. It refers to investigations or analyses carried out to improve understanding of a phenomena, and to the findings of this work. There is no assumption that research is necessarily conducted by academics or that it adheres to any particular standards. The concept of quality is considered in relation to the purpose of the research and its theoretical basis.</td>
</tr>
</tbody>
</table>
**Research use / utilisation**

SPIRIT defined research use as "the extent to which research is sought and used in developing the policy or programme document, taking into account barriers and facilitators". In this thesis research use is conceptualised more broadly as any consideration of research or research-informed information, regardless of how it was encountered or what is subsequently done with it. My aim was not to explore any definition of use, but to take a more grounded approach in investigating how policymakers perceive this use. I ask readers to forgive the term research utilisation. I do not use it to sound grand but because of the inability to distinguish in writing between use (aka utilisation - a noun) and use (a verb). This can be confusing. Also, because utilisation emphasises that something has been put to good use, which is the aim of research utilisation strategies (for example, see https://english.stackexchange.com/questions/143941/when-to-use-use-and-when-to-use-utilize-in-a-sentence).

**Research transfer, translation, exchange, or mobilisation**

See Knowledge exchange, transfer, translation or mobilisation

**SAGE**

One of the three outcome measures used to evaluate SPIRIT. SAGE was a structured interview that sought to ascertain the extent to which agency staff used research in the development of the policy and program documents. See Chapter Five.

**SEER**

One of the three outcome measures used to evaluate SPIRIT. SEER was an online self-reported survey that sought to ascertain how policymakers currently use research, their confidence, how they value research, and how supportive they feel their workplace is of research use. See Chapter Five.

---

**Acronyms and initialisms**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIPHER</td>
<td>Centre for Informing Policy in Health with Evidence from Research</td>
</tr>
<tr>
<td>CMO</td>
<td>Context+Mechanism=Outcome</td>
</tr>
<tr>
<td>LP</td>
<td>Liaison person (the internal members of staff nominated to coordinate SPIRIT in each agency)</td>
</tr>
<tr>
<td>NSW</td>
<td>New South Wales (the most populous state in Australia. Sydney is the state capital)</td>
</tr>
<tr>
<td>ORACLe</td>
<td>Organisational Research Access, Culture and Leadership</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomised controlled trial</td>
</tr>
<tr>
<td>SAGE</td>
<td>Staff Assessment of enGagement with Evidence</td>
</tr>
<tr>
<td>SEER</td>
<td>Seeking, Engaging with and Evaluating Research</td>
</tr>
<tr>
<td>SPIRIT</td>
<td>Supporting Policy In health with Research: an Intervention Trial</td>
</tr>
</tbody>
</table>
Appendix 1. references

Appendix 2. Models of research utilisation

Seven overarching models or ways conceptualising the use of research in policy are apparent in the literature [e.g. 1, 2-5]. These models envisage the problem in different ways, and so suggest different potential pathways through which research and policy interact. I will briefly outline each model and explore some of their implications for a research use intervention in policy agencies.

**Instrumental or rational.** Instrumental use of research occurs when policymakers treat findings as data that helps to answer questions about policy problems and directs action for solving them. The catalyst for instrumental use may be a pressing policy need, causing policymakers to actively seek research, or new research that is so compelling it drives a policy response. [6, 7]

The most prominent approach to investigating the instrumental use of research has been to identify barriers and facilitators. A recent review [8] supported previous work in finding that the main barriers to research use that policymakers reported were: 1. Poor availability and access to research, 2. Lack of clarity, relevance and reliability of research findings, 3. Poor timing or lack of opportunity, 4. Inadequate skills by policymakers in accessing and applying research, and 5. High costs. Efforts to address these barriers and enhance the facilitators often conceptualise the core problem as a gap between knowing and doing that can be bridged using the right combination of pull and push activities. Pull refers to knowledge users’ attempts to access research, and push refers to knowledge producers’ efforts to package and disseminate research. It is generally assumed that efforts are required on both sides. This is supported by a recent study that found intensive push efforts alone resulted in no uptake of research. [9] This gap between research and policy is not thought to be geographic but cultural, as exemplified in Caplan’s polemical 2-communities thesis that suggests researchers and policymakers inhabit different worlds with conflicting values, languages, timelines and rewards. [10] If treated as a strict dichotomy, this is crude and ignores the blurred boundaries between the two communities (for example, many policymakers have research expertise) [11], but it remains an evocative touchstone for articulating some of the core obstacles for research-informed policy. [12]

Instrumental use of research dominates popular discourse about the research-policy relationship yet it has been found to be relatively uncommon; certainly less common that the conceptual or rhetorical uses outlined below. [1, 13] This may be because it is predicated on a linear, rationalist depiction of policymaking that conceptualises research use as a formal process of information acquisition and application whereas, in reality, this process is diffuse, iterative and often unpredictable. [1, 13] As Weiss argued in 1983, instrumental use of research is rare because, from a policy perspective,

“...research does not examine all the relevant variables, research rarely fits the exact circumstances with which decisions are made, research is not ready on time for decisions, research conclusions are not clear or authoritative enough to provide trustworthy guidance, research reports do not reach the right audience, decision makers do not understand or trust research findings or understand how to interpret and apply them, the lessons from research are outweighed by competing considerations of agency self-interest and individual career advancement” [13:170]

Additional criticisms include the assumption that: 1. Access to and uptake of research is a unidirectional activity of knowledge transfer, 2. If policymakers have ready access to research they will use it, and 3. Those who know (i.e. researchers) are distinctly different from those who do (i.e. policymakers). [1, 14-17] Similar criticisms are directed at staged models of research use that identify linear steps, pathways or ‘pipelines’ through which research is translated into policy and practice. [1, 18]

**Opportunistic.** In sharp contrast to instrumental use, opportunistic models depict policymaking as chaotic, fragmented and often serendipitous. Kingdon describes policymaking as “organised anarchy” that comprises streams of problems, politics, and policy. [19] Only when these streams align do brief windows of opportunity open, allowing policymakers, advocates and experts (such as researchers) to advance their agendas. This model emphasises ambiguity and haphazard decision-making where policymakers delve into a figurative “garbage can” containing problems and potential solutions, and attempt to match them up. [19] In contrast, de Leeuw and colleagues’ stress the skill and diligence with which policymakers make sense of
and structure apparent chaos as they manage competing interests and power dynamics, strategise, plan and trade off costs and benefits. In their model information such as research is sought, but is used only if it can be synthesised into a negotiated “policy logic”. [20]

**Incremental.** Incrementalism, or “muddling through”, argues that policy evolves in small value-driven steps, allowing policymakers to adjust the direction of further reform in response to deepening understanding of how the policy in practice reflects social values and objectives. This understanding is reached through empirical evaluation and negotiation. [21] The punctuated equilibrium model builds on this to describe how incrementalism is occasionally interrupted by instability and intense policy change when external pressures reach a critical mass. [22] Incremental policymaking involves myriad players and activities, thus research is likely to interact with policy at multiple points and through diverse people and channels. [1] Incremental models emphasise how closely policy options orbit the status quo, a phenomenon reflected in the ‘least coercion rule’ where policymakers choose interventions that will have minimal impact on people’s individual choice. [20]

**Conceptual.** In conceptual models research is not sought or used intentionally. Rather, ideas from research infiltrate policymakers’ perceptions through a process of “enlightenment” and, over time, influence their beliefs and actions. [4, 6, 7] In this way research findings sensitise policymakers to new issues, define or redefine problems, and shape the policy agenda. Conceptual use is considered to be common (more so than instrumental use [1]), but it is extremely hard to assess as policymakers can seldom pinpoint specific studies or a clear influence on decisions. [4] Worryingly, conceptual research impact has no quality control, meaning that what is absorbed is often incomplete, biased or obsolete, and likely to be distorted though informal translation processes and inaccurate recall. This can result in “endarkenment” where policymakers feel informed, but are not. [13:268]

**Rhetorical or political.** Rhetorical models emphasise the pluralistic, contested nature of policymaking, which is seen as a “messy unfolding of collective action, achieved mostly through dialogue, argument, influence and conflict” [23]. Power, interests and ideology are critical mediators [24, 25], thus key concerns for policymakers are managing ambiguity, bargaining, coping with institutional constraints and reconciling competing views. [15, 26] In this environment, research is seen as value-laden, like any other knowledge claim, and it is used strategically as argument or ammunition [6] in a “formal struggle over ideas and values”. [27:45] Rhetorical use of research may involve persuading colleagues, bolstering one’s position (including the retrospective justification of a priori decisions), deflecting criticism or bludgeoning opposition. [4, 6, 7] Although such uses of research are often decried, Weiss argues that they are illegitimate only when they misrepresent findings. [4] Another variant is tactical use where claims about forthcoming research are used to defend inaction. [4] Many state that rhetorical use is the dominant mode of research use in policymaking; greater than instrumental or conceptual use, and that it is fundamental if research is to be considered within the argumentative discourse of political action. [10, 15, 27-29]

**Interactive.** Interactive models stress the centrality of interpersonal relationships for getting research into policy. They shift the emphasis from ‘push’ and ‘pull’ activities towards interpersonal communication and collaboration between policymakers, researchers, practitioners and other stakeholders [4, 30] and, in some cases, the use of knowledge brokers and other intermediaries. [31, 32] In these models trust, shared understanding, deliberation and partnership approaches are seen as the primary mechanisms. [1, 33, 34] The deliberative process increases the likelihood that policy needs inform the research design so that applicability is maximised, and that, having some ownership of the research development process, policymakers have an investment in its use. As Nutley et al. put it, “the best and most lasting influences of research come about not when information is linearly transferred to the practitioner, but when teams of practitioners and researchers co-create knowledge by working together”. [1] Interactive models leverage people’s inclination to accept information from people they know rather than from formal sources [33] and build on policymakers’ assertions that personal interaction with researchers is critical for supporting their use of research. [1, 32, 35] This is recognised by ‘advocacy coalitions’—groups with shared beliefs who strive to influence the views of key decision-makers. [36]
Research/policy interactivity can be conceptualised on a continuum of involvement from providing information, to involvement in some decisions, through to full collaboration [e.g. 37] but, as interactive approaches gain more attention, the emphasis is increasingly towards the more collaborative end of the spectrum where research is co-produced. [38] Some characterise this shift as an evolution towards Mode 2 knowledge production which, in contrast to investigator-initiated, discipline-based academic research of Mode 1, is collaborative, multidisciplinary and focused on developing problem-solving knowledge. [39] In contrast to instrumental approaches that strive to bridge boundaries between the worlds of research and policy, co-production aims “to dissolve the boundary between producers and users”. [40] Formalised research/policy partnerships and co-production have been found to increase the uptake of research and to strengthen the capacity of their members. [38, 40-44]

However, interaction can be fraught and unproductive. [45] For example, Gabbay et al.’s study of health services planning consultation showed that group decision-making was distorted by personal, professional and political agendas. [46] Elliott and Popay found that policymakers and researchers often clashed, leading them to conclude that genuine dialogue was more of an ambition than a reality. [47] Other studies suggest pervasive difficulties in negotiating partnership resources, tasks and expectations, and in producing policy-useful outputs. [48-50] There are risks, constraints and power imbalances in the “politics of co-production”. [51]

**Systems.** These models reflect an increasing interest in how research and policymaking deal with complexity, and the implications for research-informed policymaking. Systems approaches eschew the traditional focus on gaps and boundaries and, instead, focus on structure and interactions in which organisations and other social institutions are conceptualised as dynamic, self-organising networks that change more like ecologies than engineered systems. [52-56] They draw on the features of complex adaptive systems such as interdependence, emergence and non-linearity [57] to show how change in one part of a system can produce unpredictable changes in other parts. [5] Systems approaches challenge accepted ways of defining social problems and critique some conventional modes of research (particularly experimental research that attempts to control contextual variables [58, 59]). In this mode, knowledge mobilisation, and policymaking in general, are best served not by trying to control or manipulate systems, but by fostering emergent conditions through strategies such as distributed leadership, networking platforms, increased co-production, and replacing rigid plans with minimum specifications that enable local creativity. [5, 60, 61] This can be supported by decision-making that is responsive to changes in the policy environment. [30]

The seven models outlined above are not mutually exclusive. The different ways that research is used are often entangled and may occur concurrently or at different points in a policy process. [2, 6] For example, research is used rhetorically in opportunistic, incrementalist and interactive styles of policymaking. This is because policy is simultaneously technical (where research is used as a content-based rationale for action), conceptual (where research is used for ideas), political (where research is used as an instrument of value-laden persuasive performance) and relational (where research is used in stakeholder consultation and negotiation processes). [2] The extent to which any of these models is in operation will also, of course, be subject to interpretation by different stakeholders. [1]
Appendix 2. references

Appendix 3. Review characteristics and search strategies

Table 1. Features of this review in relation to characteristics of scoping and rapid realist reviews

<table>
<thead>
<tr>
<th>Elements of a review</th>
<th>Characteristics of scoping reviews [1-4]</th>
<th>Characteristics of (rapid) realist reviews [5-8]</th>
<th>Characteristics of this review</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aims</strong></td>
<td>To map and summarise a body of literature, identify useful findings, and identify research gaps. Stakeholder consultation may be used to increase the usefulness of findings</td>
<td>To advance understanding of which interventions work for whom, in what circumstances, and how. Rapid reviews engage stakeholders in the process</td>
<td>Our aims correspond with both types of review, but no stakeholders were involved as the review was to inform our own program of work</td>
</tr>
<tr>
<td><strong>Research question</strong></td>
<td>A broad question to investigate what has been done in a field</td>
<td>A realist question that tackles the issue of what works for which groups in which circumstances</td>
<td>The research question is exploratory and reasonably broad, but with a realist focus on mechanisms</td>
</tr>
<tr>
<td><strong>Search strategy</strong></td>
<td>Inclusive (e.g. often incorporates grey literature), iterative, transparent but not necessarily replicable due to use of citation snowballing etc.</td>
<td>Same as for scoping reviews but guided by an initial program theory that is refined throughout the search. Often trans-disciplinary</td>
<td>Our search strategy corresponds with that of a scoping review. No overarching program theory or causal hypothesis was used to frame the search or analysis</td>
</tr>
<tr>
<td><strong>Additional information</strong></td>
<td>Optionally, authors may be contacted for supplementary information</td>
<td>Authors may be contacted for supplementary information, but this is unlikely when the review is rapid. Explanatory theory is sought</td>
<td>Authors were not contacted. Explanatory theory was sought, but not with the dedication associated with a high-quality realist review</td>
</tr>
<tr>
<td><strong>Quality appraisal</strong></td>
<td>Quality criteria are usually either low threshold or none are used. Critical interpretive synthesis (CIS), which shares some commonalities with scoping reviews, excluded only papers that are judged to be “fatally flawed” [9]</td>
<td>Quality is assessed in relation to the review question: 1. Relevance – if it can contribute to theory building and/or testing; and 2. Rigour – if the method used to generate that data is credible and trustworthy</td>
<td>Our approach is closest to scoping reviews and CIS in that we prioritised relevance re the studies’ goals, strategies and participants rather than theory-building. Relevant studies were included providing they were not “fatally flawed” [9]</td>
</tr>
<tr>
<td><strong>Data extraction and synthesis</strong></td>
<td>A descriptive analytical approach that includes process and theoretical information</td>
<td>Focuses on demi-regularities, middle range theories, context+mechanism=outcome patterns</td>
<td>We used realist methods, following the six-stage approach described by Best et al. [10]</td>
</tr>
<tr>
<td><strong>Presentation of findings</strong></td>
<td>Depends on aims of review and types of studies included. May be tabular, narrative or both.</td>
<td>Incorporates context+mechanism=outcome patterns in tables, figures and/or propositions. Focus on theory-building and testing</td>
<td>Key findings are presented in context+mechanism=outcome configurations</td>
</tr>
<tr>
<td><strong>Philosophical underpinnings</strong></td>
<td>Unspecified</td>
<td>Realist epistemology and ontology, including assumptions about causality</td>
<td>Our philosophical stance is realist</td>
</tr>
</tbody>
</table>
### Table 2. Database search strategies

<table>
<thead>
<tr>
<th>Search strategies and rationale</th>
<th>PAIS (formally Public Affairs Information Service)</th>
<th>Web of Science (WoS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Syntax used</strong></td>
<td>ALL fields: (intervention OR program*) AND (research OR evidence OR knowledge OR data) AND (policy-makers OR &quot;policy makers&quot; OR policymakers OR &quot;government department&quot; OR (government AND decision-makers OR &quot;decision makers&quot;) OR &quot;policy agency&quot; OR &quot;policy organisation&quot;) AND (capacity OR capability OR &quot;professional development&quot; OR skills OR training OR collaborat* OR partner*)</td>
<td>TOPIC: (intervention OR program*) AND (research OR evidence OR knowledge OR data) AND policy-makers OR &quot;policy makers&quot; OR policymakers OR &quot;government department&quot; OR (government AND decision-makers OR &quot;decision makers&quot;) OR &quot;policy agency&quot; OR &quot;policy organisation&quot;) AND (capacity OR capability OR &quot;professional development&quot; OR skills OR training OR collaborat* OR partner*)</td>
</tr>
<tr>
<td><strong>Additional filters</strong></td>
<td>Language: English</td>
<td>Language: English</td>
</tr>
<tr>
<td></td>
<td>Source type: Reports, Scholarly Journals</td>
<td>Documents: Articles</td>
</tr>
<tr>
<td></td>
<td>Date range: 2001 to 2016</td>
<td>Date range: 2001-2016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Indexes: SCI-EXPANDED, SSCI, A&amp;HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Refined by: HEALTH POLICY SERVICES</td>
</tr>
<tr>
<td><strong>Articles returned</strong></td>
<td>215</td>
<td>255</td>
</tr>
<tr>
<td><strong>Rationale for selecting this database</strong></td>
<td>PAIS is considered to be the world’s largest database for public policy-related peer-review and grey literature [11]. It was selected because of its breadth in content, including government and NGO reports, which supplemented the more academic databases used by the Moore and Campbell reviews. Also for its orientation to political science - a discipline that is strangely neglected in the research utilisation literature, even where it pertains to policymaking [12, 13], Harrow [14] describes the concept of capacity building as “theoretically homeless” but best served by the literature on public administration, stewardship and community development—all of which are included in PAIS.</td>
<td>WoS was selected because it is the world’s largest collection of research across the sciences, social sciences and humanities. However, we only searched within one category—health policy services—due to the unmanageable number of returns without that filter (2311). WoS also provides excellent citation searching which was used for snowball searching.</td>
</tr>
</tbody>
</table>
Appendix 3. references

6. The RAMESES Project. *Quality standards for realist synthesis (for researchers and peer-reviewers)* 2014 (cited 2017 27 Feb); Available from: [www.ramesesproject.org](http://www.ramesesproject.org)
### Appendix 4. Overview of included studies

<table>
<thead>
<tr>
<th>Study reference</th>
<th>Study design, goals, intervention strategies and domain</th>
<th>Participants and setting</th>
<th>Evaluation methods</th>
<th>Outcomes of interest/ measures[^35]</th>
<th>Theories, models, frameworks &amp;/or theses</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study 1.</strong></td>
<td>Intervention study[^36] evaluating policymakers’ perceptions of relevance and potential impact of a long-term Policy Liaison Initiative (PLI) aimed at supporting the use of Cochrane systematic reviews in policy work. Intervention strategies included: a community of practice to increase awareness and support knowledge sharing, seminars, skills workshops, a tailored website and review summaries. Domain: Access, Skills improvement and Interaction</td>
<td>Targeted participants: Policymakers at managerial and lower levels in the federal Department of Health. Participants n=unknown Country: Australia</td>
<td>Individual interviews with participants: n=10/38 managers (who were randomised and sent personalised invitations). Seven group interviews n=33/5000 staff across all levels of the Department (who were sent general invitations). Plus participation data.</td>
<td>Use and awareness of systematic reviews Awareness and relevance of PLI Individual-, unit- and organisation-level capability to assess, interpret and apply research Links with researchers and other external experts</td>
<td>Referenced literature regarding the complexity of policymaking (e.g. [2, 3]), the need for accessible research, and the value of high quality systematic reviews as efficient decision-making aids (e.g. [4, 5]). Study outcomes were informed by arguments that distal research use cannot be wholly attributed to capacity-building (e.g. [6]). Data analysis was guided by the theoretical domains framework [7] and built on themes in previous studies (e.g. [8]).</td>
<td>Despite &gt;565 occasions of attendance at forums and 294 members, most interviewees were not aware of PLI. They used reviews/syntheses but most did not distinguish between these and systematic reviews. Some did not understand the scope of systematic reviews. Access was impeded for those who found the Cochrane Library hard to navigate. Links with researchers bolstered capacity to access and use research. Policy-relevance, applicability and accessibility were key needs. Managers were more confident than general staff that the Department had the skills to acquire, assess and interpret research.</td>
</tr>
</tbody>
</table>

[^35]: Not all of the outcomes of interest/measures are explicitly stated in the articles. Many are inferred from the descriptions of data collection and results.

[^36]: Three terms are used to describe the study design: Experimental = some form of randomisation and control groups were used, Intervention = the research team provided the intervention and evaluated it, but not using experimental methods, and Observational = the intervention or initiative being evaluated was not designed as part of a research study.
<table>
<thead>
<tr>
<th>Study reference</th>
<th>Study design, goals, intervention strategies and domain</th>
<th>Participants and setting</th>
<th>Evaluation methods</th>
<th>Outcomes of interest / measures</th>
<th>Theories, models, frameworks &amp;/or theses</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study 2.</strong> Brownson et al 2011 [9]</td>
<td>Experimental study to identify the factors that influence whether state policymakers would find evidence briefs about mammography screening understandable, credible and useful. States were stratified and participants randomised to four groups, each receiving one type of brief: 1. Data-focused brief with state-level data, 2. Data-focused with local-level data, 3. Story-focused with state-level data, 4. Story-focused with local-level data. Domain: Access</td>
<td>Three groups of state-level policymakers from six states: state legislators (elected officials), legislative staff (those serving the legislators), and health executive branch administrators (civil servants). Participants n=840 Country: USA</td>
<td>Questionnaire-based study. Post-intervention survey responses n=291, an overall response rate of 35%, but a 47% response rate from executive branch administrators (the group of interest in this review)</td>
<td>Whether the brief was understandable, credible, likely to be used, and likely to be shared</td>
<td>Described the contradictory and overwhelming volume of information policymakers receive, and their preference for concise, relevant syntheses (e.g. [10, 11]). Noted the power of narrative in policy communication and the composition of effective policy briefs (e.g. [12]). Stories were crafted as per Kreuter et al.’s [13] framework. Data collection and analysis referenced personal and professional influences on policymakers’ info engagement. [e.g. 14, 15]</td>
<td>All recipients found the briefs understandable and credible. 67% of executive policymakers reported the briefs contained an appropriate amount of information, but 20% wanted more. This group were more likely to use and to share data-focused than story-focused briefs. This was the same for legislators but not staffers who were most likely to use story-focused briefs. Participants favoured state-level rather than local data, but they all operated at state level so a regional policymaker cohort may have responded differently.</td>
</tr>
<tr>
<td><strong>Study 3.</strong> Campbell et al. 2011 [16]</td>
<td>Observational evaluation of policymakers’ satisfaction with the process and outcomes of Evidence Check, a program that helps policymakers commission high-quality rapid reviews of research in 6-8 weeks. An Evidence Check involves: policymakers completing a commissioning tool, a knowledge brokering session to clarify needs, agreement on the review proposal, selection of suitable researchers, management and delivery of the review. Domain: Access</td>
<td>Policymakers in state government who commissioned Evidence Checks during 2007-2008 Participants n=&gt;30 Country: Australia</td>
<td>Interviews with eight policymakers who had commissioned Evidence Check reviews. Independent researchers assessed the accuracy of six commissioned rapid reviews that were randomly selected.</td>
<td>Satisfaction with the knowledge brokering process Satisfaction with agreed review questions and parameters Relevance and policy impacts of the review product Relevance and accuracy of reviews</td>
<td>Like study 1 (above), barriers to research use and the value of concise syntheses were identified; but the emphasis here was on the limitation of formal systematic reviews [e.g. 17]; policymakers’ need for timely, accessible and applicable answers to specific questions; the benefit of linkage with researchers; and the use of knowledge brokers as expert boundary spanners who can facilitate communication and enable the production of better-targeted syntheses (e.g. [18-21]).</td>
<td>Participants reported high levels of satisfaction with the knowledge brokering process and the reviews produced. Knowledge brokering helped to: refine research questions, shape project parameters (e.g. scope, budget, timeframe), and facilitate communication with researchers. The reviews were seen as useful with mostly indirect impacts, e.g. informing policy deliberations and identifying evidence gaps. Independent researchers assessed the reviews as accurately reflecting the current body of evidence.</td>
</tr>
</tbody>
</table>

37 The terms evidence brief and policy brief are often used interchangeably in the literature. We use one term—evidence brief—for all studies, irrespective of the term used by each study’s authors.
<table>
<thead>
<tr>
<th>Study reference</th>
<th>Study design, goals, intervention strategies and domain</th>
<th>Participants and setting</th>
<th>Evaluation methods</th>
<th>Outcomes of interest / measures</th>
<th>Theories, models, frameworks &amp;/or theses</th>
<th>Results</th>
</tr>
</thead>
</table>
| **Study 4.**  
Dagenais et al.  
2013 [22] | Observational evaluation of a ‘knowledge transfer’ strategy aimed at improving the use of research results by policymakers and other stakeholders.  
Intervention strategies included the production and dissemination of evidence briefs, workshops for sharing evidence and experiences, forums, research dissemination, advocacy documents and communications. Nearly 50 activities.  
Domain: Access & Interaction | Policymakers from the national Ministry of Health and regions, funders, NGOs and health systems managers in two health districts  
Participants n=“hundreds of people”  
Country: Burkina Faso | Mixed methods case study. Baseline survey, participant interviews (n=38) and document review. | Extent to which aspects of the strategy were perceived as helpful  
Extent to which research findings were used | The authors note the trend towards models of knowledge translation that emphasise complexity and policymaker-researcher interactions. [23]  
Seminal research use typologies are referenced (e.g. [24, 25]). A framework of categories that support research use is identified from the researchers’ previous studies that guide the data analysis (although not cited, this framework has much in common with the diffusion of innovations framework developed by Greenhalgh et al. [26]). | Few participants read the research documents. Dissemination workshops had the greatest impact as they bypassed the need for skills in reading, appraising and interpreting research and did not contribute to information overload. However, workshop attendance was uneven. It was hard to get Ministry staff, especially those at senior levels, to participate in the intervention or evaluation and there were no discernible impacts at national level. Some regional policymakers were unaware of the research findings, but others used them instrumentally, conceptually and tactically. |
| **Study 5.**  
Dobbins et al.  
2001 [27, 28] 38 | Intervention study that tested the extent to which health decision-makers used policy relevant systematic reviews that were provided by the research team.  
24-month trial with five systematic reviews (on topics of current policy relevance) disseminated once.  
Domain: Access | ‘Public health decision-makers’ included clinicians, program directors and program managers in public health units.  
Participants n=unknown, but 41 public health units participated  
Country: Canada | Cross-sectional telephone survey with 141 of 147 invited decision-makers (96% response rate), and a self-administered organisational demographical questionnaire was completed by 35 of the 41 public health units (85%). | Extent to which policymakers used the reviews in decision-making  
What characteristics predicted use at the levels of the review, the individual policymaker, the organisation, and/or environment | This study was framed by diffusion of innovations theories (e.g. [29]). It drew links between research use in practice and policy in relation to the impact of: multiple forms of evidence, the power of personal attributes and experience, and the complex processes whereby new initiatives are adopted (e.g. [30-32]). An unpublished (and undescribed) framework guided the study. Survey instruments derived from previous studies, two of which combined concepts from multiple studies, mostly in nursing (e.g. [32-35]). | 63% of respondents said they had used at least one systematic review to make a decision. Reviews were most useful for program justification and planning, but had little impact on evaluation decisions. Predictors of use were: organisational position (managers and directors were significantly more likely to use a review than clinicians), expecting to use a review in the future, perceptions that the reviews were easy to use and compensated for limited critical appraisal skills. Their impact was rated more highly in agencies with higher existing levels of support for research use. |

---

38 The two articles cited in relation to this study as complementary articles about the same study, so aspects of both are synthesised here  
39 Public health units are municipal-level agencies with legislative responsibility for research-informed program planning and evaluation (see Kothari et al. 2005)
<table>
<thead>
<tr>
<th>Study reference</th>
<th>Study design, goals, intervention strategies and domain</th>
<th>Participants and setting</th>
<th>Evaluation methods</th>
<th>Outcomes of interest / measures</th>
<th>Theories, models, frameworks &amp;/or theses</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 6.</td>
<td>Experimental (randomised controlled) trial comparing: (a) access to an online registry of systematic reviews, or (b) access to the registry plus tailored weekly messages, or (c) access to the registry plus tailored messages and knowledge brokering. Health departments were stratified and randomly allocated to the three groups. Over the 12-month trial there was ongoing access to the registries. Tailored messages were sent weekly X7. Knowledge brokers communicated &gt;once monthly, made a site visit of 1–2 days, and hosted workshops and webinars. Domain: Access and systems improvement</td>
<td>Policymakers and program managers in regional and local public health departments. Participants n=unknown, but 108 of 141 (77%) national health departments took part</td>
<td>Telephone-administered surveys twice at baseline and once post-intervention. Questions had been previously used and tested for reliability and validity. Post-intervention data were collected from 88 of 108 (81.5%) participating public health departments.</td>
<td>Extent to which research was considered in a program-planning decision within the past 12 months Use of research in policies and programs</td>
<td>Barriers to research use were identified, including time constraints, research availability, and policymakers’ limited capacity to appraise and translate studies (e.g. [19, 33]). References the value of tailored and targeted messaging, and of knowledge brokers, to improve the use of systematic reviews (e.g. [37, 38]). The study is guided by a framework that integrates concepts from diffusion of innovations [29] about the stages of adoption of new initiatives, plus concepts from the authors own work about the characteristics that mediate the uptake of research.</td>
<td>In most policy areas, the intervention had no significant effect on evidence-informed decision-making, with no significant difference between the three intervention groups in the extent to which research was used. In public health there was a significant between-group difference in research use only when access to both systematic reviews and tailored messages were combined. [36] Having access to an online registry of research appeared to have no impact at all. Knowledge brokering also appeared to be ineffective, but may trend toward a positive effect when organisational research culture is perceived as low.</td>
</tr>
<tr>
<td>Study 7.</td>
<td>Observational study of a facilitated engagement strategy that enables researchers to present contextualised findings to policymakers. Included validation of an evaluation instrument. From 2008, 23 seminars (presentations with Q&amp;A) and 13 roundtables (tailored interactive discussions) have been held. They have been facilitated by a knowledge broker since 2011. Domain: Access and Interaction</td>
<td>Policymakers in the federal Department of Health Participants n=1865(?)</td>
<td>Post-forum questionnaires were completed, n=979, (52.5% response rate). Questions focused on effectiveness, relevance and receptivity</td>
<td>Perceived effectiveness in broadening knowledge and stimulating thinking Perceived relevance (work applicability) Research receptivity (interest in and use of research in past 12 months)</td>
<td>The study is framed by literature focusing on the complexity of getting research into policy. [40] The goal is conceptualised as research mobilisation (rather than transfer or translation [41]), and linkage between researchers and policymakers [42] that counters the two-communities divide. [43] The need for tailored information, and the situated nature of research usefulness are emphasised. [44, 45] Forums were based on exchanges in previous studies (e.g. [46])</td>
<td>Participants indicated that the forums had broadened their knowledge and stimulated thinking. Over ¾ indicated the forums’ content was directly applicable to their work and they may be able to use it. The content of roundtables was more applicable than seminars, but was no more effective in stimulating thinking and/or broadening participants’ knowledge. International speakers were rated as especially effective. Nearly 90% had used research in the past 12 months and said they would use it more if it were easily available.</td>
</tr>
</tbody>
</table>

[References: 36, 37, 38, 40, 41, 42, 43, 44, 45, 46]
<table>
<thead>
<tr>
<th>Study reference</th>
<th>Study design, goals, intervention strategies and domain</th>
<th>Participants and setting</th>
<th>Evaluation methods</th>
<th>Outcomes of interest / measures</th>
<th>Theories, models, frameworks &amp;/or theses</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 8. Hawkes et al. 2016 [47]</td>
<td>Observational study of interventions to enhance the capacity of policymakers, policy organisations and wider policy influencers. Situation analysis was used to identify capacity needs and inform intervention design. Strategies included: skills workshops, seminars, cross-sector retreats, research access infrastructure improvements (including a government-run portal for research syntheses), incorporation of research-policy topics in post-graduate health training, and establishing a government committee for commissioning reviews and advice. Domain: Access, Skills improvement, Systems improvement, Interaction</td>
<td>Policymakers in government health departments, parliamentarians, senior health care managers and practitioners – varied according to country. Participants n=unknown Countries: Bangladesh, Gambia, India and Nigeria</td>
<td>Implementing sites conducted their own evaluation. Methods included: stakeholder interviews (Bangladesh, Nigeria), pre/post quantitative surveys of changes in knowledge, attitudes and practice (India, Nigeria), and document analysis of pre/post frequencies of references to research in Parliamentary discussions (Gambia). These evaluations were synthesised by an independent evaluator.</td>
<td>Extent to which the tailored intervention strategies addressed the capacity needs (individual, organisational and institutional) that were identified locally in each setting</td>
<td>The interplay of evidence and politics is noted [48, 49], and the capabilities required to use research effectively in this complex environment [50], which are often lacking in LMIC. [51] Capacity is conceptualised at multiple levels as per the UK Department of International Development. [52] Institutional capacity is regarded as especially critical for sustained research-informed policymaking. The authors cite Ward et al. [53] on the role of interaction as an explanatory feature in research transfer models, and frame the results using categories of capacity described by Moore et al. [54]</td>
<td>Skills workshops were well-received and generally well attended, and were successful in building individual capacity to access, understand and use research/data. Interactions and provision of policy-relevant summaries led to improvements in researcher-policymaker relationships. Organisational capacity to access research was strengthened via infrastructure supports (e.g. IT resources) and newly established interactive forums. The impacts on policymaking, however, were intangible. Institutional capacity was seldom addressed and the authors conclude that more needs to be done in this sphere. Lack of shared evaluation frameworks hindered the study.</td>
</tr>
<tr>
<td>Study reference</td>
<td>Study design, goals, intervention strategies and domain</td>
<td>Participants and setting</td>
<td>Evaluation methods</td>
<td>Outcomes of interest / measures</td>
<td>Theories, models, frameworks &amp;/or theses</td>
<td>Results</td>
</tr>
<tr>
<td>----------------</td>
<td>------------------------------------------------------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>--------------------------------</td>
<td>-----------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Study 9. Hoeijmakers et al. 2013 [55]</td>
<td>Observational study of the effects of Academic Collaborative Centres (ACCs): partnerships between government, public health services and universities to support knowledge mobilisation. Activities included jointly creating a program theory, collaborative research projects, education forums, and shared data infrastructure. Domain: Access, Skills improvement, Systems improvement and Interaction</td>
<td>Policymakers and practitioners in municipal departments and public health services, and researchers in public health Participants n=unknown Country: the Netherlands</td>
<td>Three-year-long developmental evaluation using mixed methods: focus groups, interviews, network analysis and case studies.</td>
<td>Extent of outputs and outcomes as defined in program theory, e.g. collaboration on research and grants, structure for knowledge exchange, implementation capacity, research uptake and appraisal skills, new publications.</td>
<td>ACCs were conceptualised as boundary organisations [56] intended to support cross-sector collaboration that would, in turn, foster Research→Policy. [57] The study’s program theory (which was programmatic rather than theoretical) guided data collection and analysis. Findings build on previous ACC [58] including a study that took an interpretive hermeneutic approach. [59]</td>
<td>The ACC provided a platform for dialogue and interaction, but project collaborations did not extend into enduring partnerships. Most committees functioned well but thematic groups were less successful due to lack of support from managers. Overall, policymakers were less involved than researchers and practitioners. New research proposals were written but non-researcher involvement was limited and traditional research designs were used. The number of projects and participants increased over time, but the structure and density of networks was unchanged.</td>
</tr>
<tr>
<td>Study 10. Kothari et al. 2005 [60]</td>
<td>Experimental case controlled study that tested whether policymakers were more likely to use a research report if they were involved in its production. 12-month trial with ongoing feedback and one presentation for the three ‘involved’ units. Both they and the three comparison units received a copy of the final report. Domain: Access and Interaction</td>
<td>Teams of general staff and managers in public health units Participants n=unknown. Three teams received the intervention and three teams were selected as controls. Country: Canada</td>
<td>Comparative multiple-case study design using group interviews with participants, individual telephone interviews with directors, and document review</td>
<td>Extent to which participants: 1. Received reports 2. “Processed” the reports (including if they assessed their merit and validity) 3. Applied the report i.e. used it conceptually or instrumentally</td>
<td>The authors hypothesise that formal policymaker-research linkage and exchange [42] will create shared agendas, solutions, practices, lexicon and goals that bridge the two-communities [43] and counter static research transfer models. The study was guided by a conceptual model of stages and types of research use ([61] and [49]).</td>
<td>Staff within units that were involved in the production of a research report were more likely to receive a report, and to understand it better and value it more, than units that were not involved. But actual use was not affected. Both involved and comparison units used the research findings to confirm that their program activities were consistent with evidence, and to compare their program performance relative to other units.</td>
</tr>
<tr>
<td>Study reference</td>
<td>Study design, goals, intervention strategies and domain</td>
<td>Participants and setting</td>
<td>Evaluation methods</td>
<td>Outcomes of interest / measures &amp;/or theses</td>
<td>Theories, models, frameworks &amp;/or theses</td>
<td>Results</td>
</tr>
<tr>
<td>----------------</td>
<td>------------------------------------------------------</td>
<td>-------------------------</td>
<td>-------------------</td>
<td>---------------------------------------------</td>
<td>------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Study 11. Kothari et al. 2014 [62]</td>
<td>Observational study to determine the extent to which the PreVAiL (Preventing Violence Across the Lifespan) research network built effective partnerships among network members. PreVAiL is an international interdisciplinary public health network comprising researchers and policy/advocacy partners. Strategies included: collective (Delphi process) research priority-setting, funded team meetings, seed grants for collaborative research and dissemination, and capacity development for early career researchers. Domain: Interaction</td>
<td>Researchers (including trainees) and knowledge users (including policymakers in different parts of public health such as justice and child welfare) Participants n=60 + 15 trainees Country: Canada, but with partners in US, UK, Asia, Europe and Australia</td>
<td>Partnership Indicators Questionnaire (PIQ) completed by 36 PreVAiL members (n=26 researchers and n=9 partners) with a 65% response rate. 19 semi-structured telephone interviews conducted two years after network became operational.</td>
<td>Partnership quality within the network: levels of partner involvement, quality of communication, perceived value of network Initial impacts of the partnerships on the application of knowledge to policy and practice (instrumental, conceptual)</td>
<td>Gaps between knowledge and practice [63] are tied to disconnects between researchers and knowledge-users. [43] Collaboration in research development and dissemination is described and advocated for. [64-67] The authors argue that, despite recent reviews [68, 69], collaboration remains a ‘black box’ and greater understanding of partnerships is needed [70]. The PreVAiL network was based on public health approaches to violence. Collaborative development of the questionnaire and indicators is described elsewhere [45]. Thematic data analysis was used. [71]</td>
<td>Participation rates varied from 11-79%. The network was seen as beneficial for individuals and organisations. 75% of PIQ respondents felt their contributions were valued. Partners used the network as a source of synthesised information, but tended to contact the same researchers. Some partners functioned as an ‘information conduit’ to their own organisation. There were collaborations in writing papers, grants and speaking at events, but desire for greater collaboration on grants, research proposals and advocacy. Most knowledge was used conceptually, but there were examples of instrumental use. Not all policymakers felt there was a common language between network members.</td>
</tr>
<tr>
<td>Study reference</td>
<td>Study design, goals, intervention strategies and domain</td>
<td>Participants and setting</td>
<td>Evaluation methods</td>
<td>Outcomes of interest / measures</td>
<td>Theories, models, frameworks &amp;/or theses</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------------------------------</td>
<td>--------------------------</td>
<td>--------------------</td>
<td>--------------------------------</td>
<td>----------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>Study 12.</strong>&lt;br&gt;Langlois et al. 2016 [72]**&lt;br&gt;<em>(This paper reports on two studies. We only include the second one here, ‘Policy BUDDIES’, as the first does not meet our inclusion criteria)</em></td>
<td>Intervention study of a pilot initiative designed to build the capacity of policymakers to demand and use systematic reviews. Strategies included: baseline situational analysis, skills workshops, the allocation of researcher ‘buddies’ to work with policymakers on refining research questions for reviews, and an online support system for buddies. Domain: Skills improvement and Interaction</td>
<td>Policymakers in provincial government involved in programs related to Millennium Development Goals, and their researcher ‘buddies’&lt;br&gt;Participants n=unknown&lt;br&gt;Countries: Cameroon and South Africa</td>
<td>The realist mixed methods evaluation included document review (e.g. technical reports, policy documents and news media), in-depth interviews with policymakers and project staff, and a focus group with researcher buddies. A previous descriptive study explored policy contexts and research needs. [73]</td>
<td>Identify any lessons learnt re the process and impact of the Policy BUDDIES strategy</td>
<td>Focuses on the need for demand-driven research [74], and the importance of organisational culture in fostering research use. [75] Intervention design was based on studies highlighting the centrality of partnerships approaches and trusting relationships alongside the need to ensure that used research is robust and valid e.g. [76]. Data collection drew on Walt and Gilson’s policy analysis framework. [77]</td>
<td>Buddying helped policymakers to value and use research evidence, but also built the capacity of researchers to understand policy needs and provide useful support. Buddies were perceived as more objective than other experts. Interactions were necessarily iterative and required equality and trust. Institutional support and incentives for using research were important barriers/facilitators to policymakers’ involvement in generating and using evidence. Champions drove policymakers’ ownership of the initiative.</td>
</tr>
<tr>
<td><strong>Study 13.</strong>&lt;br&gt;Pappaoanou et al. 2003 [78]</td>
<td>Intervention study that tested an intervention for strengthen the capacity of policy staff to collect, analyse, report and use epidemiological data. Strategies included the implementation of country-specific health information and communication systems, extensive skills training tailored for the different participant groups and mentorship in applying those skills. Domain: Access, Skills improvement, Systems improvement</td>
<td>Policy decision-makers, program managers, technical experts, and information specialists in Ministries of Health&lt;br&gt;Participants n=unknown&lt;br&gt;Countries: Bolivia, Cameroon, Mexico and the Philippines</td>
<td>Data availability and use was measured at baseline, midpoint and 1-year after completion. Indicators were matched to country situations and project designs. Additional methods included participant interviews and country case studies.</td>
<td>Outcome indicators were matched to country situations and project designs. Examples include proportions of participants who:&lt;br&gt;1. Satisfied minimum skills requirements post-training&lt;br&gt;2. Presented their work 3. Showed improvement on test scores&lt;br&gt;4. Made data-based decisions</td>
<td>Draws on research utilisation literature re the limitations of rational research use models, the influence of political context [79], and the need to involve users in systems design. [80] Aimed to reduce barriers including: the failure of researchers to produce quality, timely, inaccessible research and lack of participation in interpretation [81]; poor systems for accessing policy-relevant information; and the need for policymakers to understand and trust health data. [82]</td>
<td>All countries trained policy staff (a) to use data and (b) to train others to use it. Participants reported the training taught them how to work as part of a public health team, empowered them to use data to identify critical health community problems, helped them understand their local decision-making environment, and helped them set achievable outcome-oriented goals and formulate and implement plans to tackle them. Quantitative skills assessment data is not reported. The intervention was found to improve data-informed public health in all countries. Some country-specific impacts are identified.</td>
</tr>
<tr>
<td>Study reference</td>
<td>Study design, goals, intervention strategies and domain</td>
<td>Participants and setting</td>
<td>Evaluation methods</td>
<td>Outcomes of interest / measures</td>
<td>Theories, models, frameworks &amp;/or theses</td>
<td>Results</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------------------------------------------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>-------------------------------</td>
<td>--------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>Study 14.</strong></td>
<td>Observational evaluation of the implementation and impacts of a strategic plan for using research in decision-making. Strategies included: skills workshops; developing literature review methods and tools; forums for sharing knowledge; restructuring and expanding the library; creating and supplementing research-related positions; accessing external expertise; and commissioning literature reviews. Domain: Access, Skills improvement, Systems improvement</td>
<td>Medical officers, library staff, directors, managers, supervisors, and research and policy analysts in a Public Health Unit. Participants n=unknown Country: Canada</td>
<td>Longitudinal qualitative case study using data from two sets of purposively sampled semi-structured interviews (n=6) and focus groups (n=27) with 70 members of the health unit, and review of 137 documents.</td>
<td>The evaluation attempted to identify: 1. How capacity change was attempted in the implementation of the strategic plan 2. Practices and resources needed to carry out research use tasks 3. Incorporation of research in decisions and documents 4. Any influencers on the above Rates of course completion Trainees’ regard for the course Trainees’ application of knowledge and skills learned during the course The contribution of this course to increasing skilled public health capacity in Ethiopia</td>
<td>The intervention was developed using strategies from outside healthcare identified in an earlier study based on the hypotheses that research-informed policymaking requires a culture of critical inquiry, staff capacity and tools for research use, and improved organisational knowledge management. Collaborative data collection and analysis was informed by key texts in organisational change (e.g. [26, 29, 85, 86]), knowledge exchange [23] and implementation [87]).</td>
<td>Over two years, staff confidence and skills increased, their literature reviews became more rigorous, research skills were built into job descriptions and evident in the changed workforce, and there was significant investment in further development. Critical factors for building capacity were identified as: strong continuous leadership; clear vision, workforce and skills development; improved access to research; fiscal investments; use of technology; better knowledge management; effective communication; receptive organisational culture and use of change management techniques</td>
</tr>
<tr>
<td><strong>Study 15.</strong></td>
<td>Intervention evaluation of a year-long program to build the capacity of government decision-makers to use HIV data strategically. The intervention comprised block weeks of training in: HIV interventions and situational analysis; descriptive and analytic epidemiology; HIV surveillance; and evaluation. Regional teams were mentored by researchers to complete a practical project that they presented for assessment. Domain: Skills improvement</td>
<td>Government employees (surveillance officers, public health laboratory technician, and project managers) in HIV/AIDS Prevention and Control Department Participants n=23 Country: Ethiopia</td>
<td>Per-module and whole-course surveys, and post intervention focus groups with trainees (n=15) and stakeholders (n=6). Participants who withdrew were followed up. Group presentations were assessed by an expert panel.</td>
<td>Training modules derived from the Centers for Disease Control and Prevention which has 50+ years’ experience of running successful training based on a “learning while doing” philosophy to build critical reasoning skills. An adapted evaluation model was used that sought to identify mechanisms that enable participants to use data in decision-making. The original model posits that training can change behaviour by influencing beliefs about the value of practices, and by increasing knowledge and skills, thereby enhancing self-efficacy.</td>
<td>92% of participants felt the course met their expectations and all said it was relevant to their work. Self-reported skills improved: trainees could collect, analyse and interpret data effectively and use the findings, and carry out work tasks confidently. The expert panel judged that trainees had learned core skills in using data but needed to refine their analyses and correct some errors. Some trainees went on to train their colleagues. Retention increased in subsequent cohorts (from 65% to 87% &amp; 92%) after program improvements, e.g. a shorter more intensive course and the addition of mentors from outside academia. Mentorship was hypothesised to be the critical mechanism of change. Stakeholders agreed the course contributed to skills capacity in Ethiopia.</td>
<td></td>
</tr>
<tr>
<td>Study reference</td>
<td>Study design, goals, intervention strategies and domain</td>
<td>Participants and setting</td>
<td>Evaluation methods</td>
<td>Outcomes of interest / measures</td>
<td>Theories, models, frameworks &amp;/or theses</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------------------------------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>-------------------------------</td>
<td>------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Study 16. Shroff et al. 2015 [91]</td>
<td>Observational evaluation of five projects resulting from a WHO initiative to “catalyse” the use of health research in policy via push, pull and exchange activities. Interventions were specified locally but could include: platforms to produce and communicate research; training programs; establishing data usage units within ministries of health; developing and using evidence briefs; and hosting policy dialogues or other forums for connecting researchers and policymakers. Domain: Access, Skills improvement, Systems improvement, Interaction</td>
<td>Participants included national and regional policymakers, and managers of healthcare institutions</td>
<td>Interviews (n=22), an evidence briefs survey (n=167, 63% response rate), a policy dialogues survey (n=140, 60% response rate), and outcome evaluation surveys (n=66 in Nigeria, n=48 in Zambia. Review of project technical reports. Not all data sources were available for all countries.</td>
<td>To understand why interventions in some settings were perceived by the key stakeholders to have made progress towards their goals, whereas others were perceived to have made little progress.</td>
<td>Referenced literature re the complexity of policymaking, different models for mobilising research, and how research is used (e.g. [8, 92, 93]). Choice of intervention strategies (evidence briefs, research packaging, policy dialogues and reflective forums) was based on previous studies (e.g. [94-96]). Jacobson et al.’s [97] theory of knowledge translation guided the identification of variables across five domains: user group, issue under consideration, research attributes, researcher-user relationships, and dissemination strategies.</td>
<td>There was considerable variation in intervention activities and their intensity. In more successful projects the use of research was aided by a combination of: enthusiastic policymakers in research-oriented ministries; research topics that were policy priorities and also interested the researchers; the availability of reliable, easy-to-understand research; positive research/policy relationships; clear expected outcomes; thorough dissemination of findings; and strong project leadership. The use of multiple strategies targeting different domains was thought to be beneficial. The practice of establishing research centres in ministries is suggested.</td>
</tr>
<tr>
<td>Study 17. Traynor et al. 2014 [98] 40</td>
<td>Observational, mostly qualitative study of the role of knowledge brokers (KBs) in policy agencies conducted across two primary studies: Dobbins et al. 2009 (included above) and the unpublished (?) case study of a subsequent trial. In the first study KBs were used as one strategy within a multi-strand RCT. In the second 22-month trial, KBs offered tailored services including group training, tools, management consultation, and intensive mentoring. Domain: Access, Skills improvement, Systems improvement</td>
<td>Policymakers in regional health departments</td>
<td>Study 1. See Dobbins et al. 2009 above. Study 2 included social network data, close-ended surveys, interviews (n=37), organisational documents and reflective journals. Thematic coding based on frequency and emphasis of themes across data sources and studies.</td>
<td>The impact of knowledge brokers in two intervention strategies</td>
<td>Knowledge brokers with requisite skills are hypothesised to use interpersonal contact to build bridges across the research/policy divide and facilitate the development and/or uptake of policy relevant research (e.g. [99-101]). Data collection and analysis was informed by principles from case study [102] and qualitative research [103, 104] texts. Findings build on previous studies to suggest that KBs enhance individual as well organisational and cross-organisational capacity.</td>
<td>KBs were found to enhance individuals’ capacity by improving knowledge, skill and confidence in searching for, appraising and applying research. Ongoing personal support throughout projects was more helpful than training or tools. Organisations used KBs to initiate train-the-trainer KB functions and research-oriented internal policies. Effective KB attributes included expertise in research and health, and personal traits of approachability and patience. Staff felt they could admit needing help because the KBs were ‘external’. Use of KBs improved as relationships grew. KBs may require organisational support to be most effective.</td>
</tr>
</tbody>
</table>

40 This article examines two studies, one of which is included in this review (Dobbins et al. 2009). However, the second study has not been published elsewhere as far as we can tell, so the article was eligible for inclusion as a primary study.
<table>
<thead>
<tr>
<th>Study reference</th>
<th>Study design, goals, intervention strategies and domain</th>
<th>Participants and setting</th>
<th>Evaluation methods</th>
<th>Outcomes of interest / measures</th>
<th>Theories, models, frameworks &amp;/or theses</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 18. Uneke et al. 2012 [105]</td>
<td>Intervention study of a program aimed at improving participants’ capacity to acquire, assess, adapt &amp; apply research in policymaking, and enhancing research/policy partnership links. The intervention comprised a one-day evidence-to-policy training forum. There were four lectures of 30 minutes given by senior academics followed by interactive sessions. Domain: Skills improvement.</td>
<td>Policymakers (directors, project/program managers, department heads in the regional Ministry of Health and uniformed services), other health decision-makers (e.g. directors of NGOs, hospital administrators) and researchers. Participants n=104 Country: Nigeria</td>
<td>Pre- and post-forum self-rated questionnaires. Six focus groups (n=9-12 people in each) were used to identify barriers to research-informed policy development.</td>
<td>Attendance rates Improvements in participants’ knowledge compared to their pre-forum status Improvements in participants’ use of evidence compared to their pre-forum status.</td>
<td>Identifies the gaps between research and policy [106] lack of health agenda ownership by policymakers in low- and middle-income countries (LMICs) [107] and the non-linear, non-rational process of policymaking. [108] The use and design of workshops was founded on work by Poulos et al. [109] Statistical data analysis used methods developed for LMICs [110] and phenomenological analysis of focus group data followed Giorgi. [111]</td>
<td>Attendance rate of 84%. Of the 81 participants 64% were policymakers. Pre-forum these participants supported the intervention goals, and believed that research could provide sound and relevant guidance for more effective, efficient and sustainable health systems. Post-forum, participants reported they had greater understanding of how to: access research and assess its policy relevance; synthesise and present research; transform research into policy; and amplify the impact of research in policy.</td>
</tr>
<tr>
<td>Study 19. Uneke et al. 2015a [112]</td>
<td>Intervention study aimed at improving participants’ capacity to develop evidence briefs. This study extends Uneke et al. 2012 (study 17 in this table). Here they expand on their previous use of a one-day training forum to include two mentored group meetings aimed at supporting policymakers to identify and investigate the evidence for potential policy options. Research syntheses were developed for each option and those assessed as having the strongest evidence base were included in evidence briefs. Domain: Skills improvement.</td>
<td>State-level health policymakers involved in the control of infectious diseases. Forum participants n=43/50 Mentored participants n=unknown Country: Nigeria</td>
<td>A pre/post survey was administered for training forum participants (n=38, but only 21 appeared to be government policymakers). The process and outputs of mentored group meetings are described</td>
<td>Changes in participants’ perceived knowledge of forum topics (collaboration, evidence briefs, policy dialogues, research ethics, and the local health policy context) Capacity to produce evidence briefs</td>
<td>Poor use of research in LMIC is noted, and the need to strengthen policymakers’ capacity to use research, including through evidence briefs. [5] They recognise the politicisation of policymaking and the need to incorporate different stakeholder perspectives in policy options. [113] Intervention design draws on studies that emphasise the benefits of training workshops and mentoring. [114] The introduction of formal group mentoring was based on guidance by the Canadian Coalition for Global Health Research. [115]</td>
<td>Results showed improvement in participants’ capacity to use research effectively in the development of evidence briefs. The self-reported post-workshop percentage increase in mean knowledge and capacity across the 5 workshop topics ranged from 21% to 46%. Mentored groups successfully produced briefs with research-informed options. The authors conclude that policymakers’ knowledge and capacity to develop evidence briefs can be enhanced via a one-day training workshop followed by an intensive mentorship program.</td>
</tr>
<tr>
<td>Study reference</td>
<td>Study design, goals, intervention strategies and domain</td>
<td>Participants and setting</td>
<td>Evaluation methods</td>
<td>Outcomes of interest / measures</td>
<td>Theories, models, frameworks &amp;/or theses</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------------------------------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>--------------------------------</td>
<td>----------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>Study 20.</strong> Uneke et al. 2015b [116]</td>
<td>Observational evaluation of a program aimed at building the capacity of a newly formed health policy advisory committee (HPAC) to promote research-informed policymaking and function as a knowledge translation platform. Strategies included: a one-day interactive workshop on evidence briefs/dialogues and use of SUPPORT tools, a three-month university-based training and mentorship program focusing on research use in health policy and systems, collaborative development of an evidence brief and hosting of a policy dialogue. Domain: Skills improvement, Interaction.</td>
<td>Directors from the regional Ministry of Health (MoH) and representatives from other government organisations, an NGO director, and senior researchers. HPAC members n=18, including n=9 MoH directors and n=4 other policy representatives</td>
<td>Interviews with HPAC members who were: 1. Workshop participants (n=8), 2. Training and mentorship program participants (n=8), and policy dialogue participants (n=8). Given that the HPAC had cross-sector membership, the number of policymakers within each sample is unknown.</td>
<td>Participants' post-training understanding of knowledge translation. Capacity to use research in practice. Quality of relationships between policymakers and researchers.</td>
<td>The intervention is modelled on HPACs in other countries and premised on the assertion that regular interaction between policymakers and researchers can address the gaps between them. [113, 117] Follows Choi et al. [118] who suggest that collaboration can increase policymakers' capacity to apply a &quot;science lens&quot; to policymaking, and researchers' capacity to be &quot;policy sensitive&quot;. Aims for a systematic and transparent appraisal of research within policy processes [5]. The evaluation drew on qualitative methods within a case study approach. [102, 119]</td>
<td>Participants reported: increased understanding of practical knowledge translation, including how to access and use research; markedly reduced distrust between policymakers and researchers; and greater ability to promote research-informed policymaking within the Ministry of Health. The evidence brief produced by the HPAC has been published [120] and is under consideration by the MoH. HPAC members called for performance measurements and institutional support to ensure continuation and independence. The authors note the need for continual training and interaction if HPAC productivity is to be sustained. 55% of participants completed the 12-18-month intervention, 63% of these produced one or more briefs (n=20) and 5 organisations developed templates for constructing future briefs. The knowledge brokering team spent an average of 30 hours per participant. Organisations with higher levels of internal support for using research developed more briefs. The program’s success was built on partnership with high-level policy staff in each organisation that were formalised and resulted in strong organisational commitment to the project, but it was undermined by high staff turnover.</td>
</tr>
<tr>
<td><strong>Study 21.</strong> Waqa et al. 2013 [121]</td>
<td>Observational process evaluation of a tailored intervention to build policymakers’ capacity to produce evidence briefs.</td>
<td>Four government agencies and two NGOs assessed as having the potential to make or influence health policies across diverse population groups and settings. Participants n=49 in 6 organisations (5-12 in each). Four of these were policy organisations</td>
<td>Intervention activities were recorded using a data collection proforma that captured information on: implementation processes and the scale, duration, reach and frequency of activities. Process diaries were kept by project team members.</td>
<td>Duration, frequency and type of interaction and/or activity between the knowledge brokering team and participants. Increase in perceived skills to acquire, assess, adapt and apply research. Application of skills in producing an evidence brief.</td>
<td>The authors hypothesised that increased researcher-policymaker interactions, facilitated by knowledge brokers [36], promote research use in policymaking. [8, 26] Intervention strategies were informed by previous studies, including: the use of an advisory panel, gaining high level organisational buy-in via ‘concept papers’ [122], targeted skills development [83], supported development and presentation of evidence briefs [123], assessment of existing skills and support for using research. [124] Use of process diaries was based on Waters et al. [125]</td>
<td></td>
</tr>
<tr>
<td>Study reference</td>
<td>Study design, goals, intervention strategies and domain</td>
<td>Participants and setting</td>
<td>Evaluation methods</td>
<td>Outcomes of interest / measures</td>
<td>Theories, models, frameworks &amp;/or theses</td>
<td>Results</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------------------------------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>-------------------------------</td>
<td>----------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>Study 22.</strong> Wehrens et al. 2010 [126] 41</td>
<td>Observational study of how interactions within a 'Small But Beautiful' research-policy-practice partnership project contributed to outcomes</td>
<td>Participants included policymakers, managers, and epidemiologists from the city council, regional health services and university staff. Participants n=unknown. Country: the Netherlands</td>
<td>Case study informed by interviews with participants (n=16), project document review and observation of project meetings and seminars.</td>
<td>The nature of relationships within the partnership. How the partnership structured and responded to research/policy expectations and interactions. How the above impacted on content emerging from the partnership.</td>
<td>This study referenced the trend away from linear research transfer and towards researcher-policymaker interaction. [8, 42, 127, 128] The lack of knowledge about how interaction contributes to research use is noted. [64, 127] It questions the hypothesis that structural support for interaction is sufficient to facilitate meaningful communication and connection. [129] Information about the intervention design is not available in English. No literature is cited as informing the data collection or analysis.</td>
<td>Goals were undermined by differences between partners in views, values and expectations. In the first presentation, policymakers perceived the researchers as poor communicators who were too focused on methodology, and found the results inaccessible and lacking policy usefulness. Results were repackaged using scenarios to highlight policy relevance, and a carefully managed public forum was held. This was perceived as successful in presenting scientifically robust findings that were also accessible and applicable. Findings have influenced problem definition and agenda-setting, and paved the way for further research-policy collaborations.</td>
</tr>
</tbody>
</table>

41 This study is based on the same initiative as that of Hoeijmakers et al. 2013 (study no. 9 above) but was conducted at a different time point by different researchers and focuses on one regional project. Hoeijmakers et al. 2013 evaluate the intervention at a national level, hence they are treated separately.
Appendix 4. references


46. Shakxson, L., et al., *Expanding our understanding of K* (Kt, KE, Ktt, KMb, KB, KM, etc.). A concept paper emerging from the K* conference held in Hamilton, Ontario, Canada, April 2012. 2012, United Nations University.


Appendix 5. Context and process

It is now widely recognised that “context is all important” when it comes to designing, tailoring, implementing and evaluating interventions. [1] Context is not noise, a background, a variable or something that happens around the edges of the intervention, but an integral part of how the intervention is implemented [2, 3], affecting its functioning in “often complex, idiosyncratic, and subtle” ways. [4] In realist evaluation, context is part of the ‘causal equation’ which determines how mechanisms operate [5], while constructionists view social structures, objects and culture as intrinsic to how people make meaning. [6] Consequently, there are increasing calls for context to be adequately investigated and reported in intervention trials to ensure that vital explanatory information is not lost. [4, 7]

Context has famously been defined as “everything that is not the intervention itself” [7], but narrower parameters are needed for the purposes of targeting process evaluation efforts. Most intervention science regards context as a combination of meso and macro features of the intervention setting, e.g. “Contexts are the physical, organisational, institutional, and legislative structures that enable and constrain, and resource and realize, people and procedures”. [8] However, realists define context somewhat differently, “Context must not be confused with locality. Depending on the nature of the intervention, what is contextually significant may not only relate to place but also to systems of interpersonal and social relationships, and even to biology, technology, economic conditions and so on” [9:6]. So, for realists, individuals’ attributes (including their reasoning and resources) and interactions are an essential part of context.

Intervention contexts are not static, therefore it is crucial that process is recognised as a key dimension of context. Organisational systems—relationships, practices, “processes, habit, and traditions” [7:557] and day-to-day circumstances—are in flux and will change throughout implementation [4]; especially in policy agencies which are highly responsive to shifts in the political environment. [10, 11] From a systems perspective, the various elements of context interact causally, leading to unpredictable changes, so that a strategy targeting one element of an organisation may ripple through the organisation in unforeseen ways. [12]

However, contextual uniqueness and unpredictability can be overstated which may prevent us from identifying important lessons about how change occurs. Indeed, some of the literature on complexity exacerbates the challenges rather than solving them [13] which can result in “...paralysing uncertainty about the unpredictable consequences of intervening in complex systems”. [14] So it is vital that we stress the importance of context, but equally that we do “not view initiatives as so constrained by context that we do not see any relevance beyond them”. [15:15] Local circumstances will always differ, but we can identify patterns (or ‘demi-regularities’ in realist parlance) that can be built on cumulatively in subsequent studies and which can provide valuable transferable lessons. As Pawson argues, “evaluation science assumes that there will be some pattern to success and failure across interventions, and that we can build a model to explain it.” [13]
Appendix 5. references

Appendix 6. Early interview questions (general participants)

Introduction
This interview is part of a larger evaluation of the SPIRIT study. My goal today is to understand as much as I can about how policy/program work is done here and how the culture of the organisation and the context you work in affect that. This will help us to assess how different aspects of the SPIRIT program work (or don’t work) in different contexts.

Work culture and context
1. I’d like to get a sense of what you do day-to-day. Can you briefly describe the work you do, focusing on your main responsibilities?
2. I’m interested in how people’s work agendas are set. Thinking about the work that you and your immediate colleagues do, who or what has the most influence in determining what gets prioritised each day?
3. How would you describe the culture of this organisation?
4. What significant changes have taken place in your workplace over the past year or so?
   • Have they/how have they impacted on you and your colleagues’?
   • How do you feel the organisation has coped/is coping with these changes?

The role of research and other information resources
5. I’m interested in the sort of information that informs your work. I have some cards here with different types of information on them.* Can you tell me which of these you use to answer the sorts of questions you deal with in your program/policy work?
6. I’d like to know about the usefulness of these types of information. Can tell me how each of these are useful to you?
   • What is this useful for? What affects how you use it?
7. How do you get hold of these types of information?
8. What makes it harder or easier to use these types of information?
9. How important is the trustworthiness (credibility, reliability) of these types of information?
   • How do you judge whether they are trustworthy or not?
10. [if research not selected: sum up overview of how and why selected information types are used, suggest why research was not included and ask if there are any other reasons]
11. [only if the interviewee uses some form of research...] You’ve described how you see the value of research in the mix of information you use. To what extent do you think your views are shared by your colleagues and by managers in this organisation?
   • Where do you think these views come from?
   • Are there any examples?
12. The way that people use information in their work can be influenced by many things: their personal knowledge, skills and background, by their work role and responsibilities, and by the expectations and culture of the organisation they work in. What do you think particularly influences the way that you use [research and other types of] information in your work?
13. I’m interested in how organisations support staff to use research. What do you see as the strengths and weaknesses in how this organisation supports staff to use research?
   • Are there any people here who particularly support or advocate for using research in policy/program work?
Before we finish, I’d like to talk about the SPIRIT program. You’ve probably only encountered one or two parts of this so far—you may have completed an online survey and possibly another interview, and you may have attended an information or feedback session. I’d like to ask you a couple of questions about your impression of the program at this early stage.

14. If you had 10 minutes with the team who designed and are implementing the program, what advice would you give them?
15. Based on your impression of the program so far, what is your prediction about how people will respond to it? Do you think it has the potential to change the way that research is used here?
16. Do you think it may have anything to offer you?
17. Given that my job is to evaluate how the SPIRIT program works in real world contexts, is there anything else I should know that would help me understand how it may or may not work in this organisation?

I would love to talk to you again in about 10 months’ time after the program is over. May I have your permission to contact you then to see if you would be willing to do a follow-up interview?

* The types of information on the cards were:

- Internal expertise/advice from colleagues
- Advice from researchers/academics
- Advice or reports from NGOs or professional bodies
- Advice or reports from private industry stakeholders
- Government reports
  - federal government
  - other states & territories
  - international jurisdictions
- Legislation
- Research commissioned by my organisation
- Research conducted by my organisation (independently or in partnership)
- Peer-reviewed research papers
- Research syntheses/systematic reviews
- Current guidelines or protocols
- Internal strategic plans/priority documents
- Internal progress reports and evaluations
- Feedback from patient/service user groups
- Consultation feedback
- Task forces/committees/advisory groups
- Data from authoritative websites
- News media articles or reports
- Public opinion polling data
- Social media (e.g. Twitter, blogs)
Appendix 7. Post-intervention interview questions (general participants)

Introduction
Before we kick off, I’d like to give you a bit of context for our conversation. We’re going to talk about the SPIRIT study. This study has three parts: 1. A program of workshops and resources that aim to improve how research is used in policy and program development, 2. Then there are three ways of measuring impacts of the program—an online survey and two types of interviews, these are being implemented every six months, and lastly, 3. A process evaluation that is exploring how the program worked (or didn’t work) in each organisation. The last one is my job and that’s what we’re doing today.

1. I’d like to start by asking why you decided to participate in SPIRIT.

2. What were your general impressions of the program?

3. Which aspects were most positive?

4. Which were most negative?

5. Do you think SPIRIT had any impact on how research is used or perceived or supported in this organisation?

   Yes

   Can you describe the impact?

   What was it about the program that enabled it to have an impact?

   What improvements would have helped re:
   - Program content
   - Program goals
   - Delivery strategies
   - Role of leadership

   No

   Why do you think that was?

   What would have needed to be different for it to have had an impact?

6. Did the program have any effect on you?

   Yes

   Can you describe the effect?

   What was it that made the difference?

   What would have needed to be different for it to have had an impact?

   No

   Why do you think that was?

   What would have needed to be different for it to have had an impact?

Have you noticed any changes in:
- How senior managers value or express value for R?
- Tools & systems that support R use?
- Colleagues’ confidence/values towards using R?
- Colleagues’ skills/knowledge in using R?

- Strategies/essential elements?
- Mechanisms? (How did it do that?)

Do you think there have been any changes in:
- The way you think or talk about the value of R in your work?
- Your knowledge, skills, or confidence in using R?

- Strategies/essential elements?
- Mechanisms? (How did it do that?)
7. Were you aware of any changes in your views or attitudes towards the program over its duration?

8. How do you think your colleagues viewed SPIRIT? What differences might there be between your views and theirs?

9. [If not already addressed...] One of the aims of SPIRIT was to engage well with the people the program was offered to. To what extent did SPIRIT manage to engage staff here? What could have been done better?

10. [If not already addressed...] It's probably safe to say that SPIRIT didn't manage to engage everyone who was eligible to participate. Without naming anyone, can you think of any colleagues who, as far as you know, chose not to participate or resisted it? I appreciate that this may be speculative, but can you hazard any explanation for this?

11. The people who took on the role of facilitating SPIRIT in each organisation were quite diverse. In your organisation X acted in that role. Do you think her/his position here or the way she/he approached the tasks involved in facilitating SPIRIT might have affected how people engaged with it?

12. Assuming there's always room for improvement, what advice would you give the people who designed and implemented the program about how to make it effective in agencies like this one?

13. I'd like you to erase SPIRIT from your mind now and imagine that you have been given the task of increasing the use of research in this organisation's work. If there were no financial, structural or political restrictions, what would you do?

14. Imagining the same idealised world with no financial, structural or political restrictions, but in this scenario you have been given the task of improving the quality of policies/programs produced by this organisation in whatever way you see fit. What would you do?

15. Lastly, given that I'm trying to figure out how and why aspects of the program worked (or didn't work) in different contexts, is there anything else it would be helpful for me to know?
Appendix 8. Headings used in the Framework Analysis matrices

**Matrix 1. Context and culture**

A : Role and tasks  
B : People and predictions re SPIRIT  
C : Agenda-setting & work processes  
D : What is it like to work here?  
E : Communication & collaboration  
F : External relations  
G : Organisational change, past and present  
H : Accessing evidence/research  
I : Valuing & appraising evidence/research  
J : Generating evidence/research  
K : Using (and not using) evidence/research  
L : Challenges/Barriers  
M : Supports/Facilitators  
N : Organisational systems and structures re research use  
O : Other drivers  
P : Anything else  

**Matrix 2. Interaction and impact**

A : Participation (what and why?)  
B : Predictions vs what happened  
C : Positive perceptions  
D : Negative perceptions  
E : Communication & understanding  
F : Value, benefit and burden  
G : SPIRIT personnel  
H : Liaison person role  
I : Other evaluative concepts (e.g. leadership, other context)  
J : Program improvement advice  
K : Change  
L : Attitudes and values  
M : Knowledge, skills & awareness  
N : Research behaviours  
O : Organisational support – policies, systems, resources, training & leadership  
P : Research relationships  
Q : Other effects  
R : Improving research use – what do you need?  
S : Improving policies & programs – what do you need?
Appendix 9. Chapter Six manuscript

Developing definitions for a knowledge exchange intervention in health policy and program agencies: reflections on process and value
Developing definitions for a knowledge exchange intervention in health policy and program agencies: reflections on process and value

Abby Haynes\textsuperscript{a*}, Tari Turner\textsuperscript{b, c}, Sally Redman\textsuperscript{a}, Andrew J. Milat\textsuperscript{d} and Gabriel Moore\textsuperscript{a}

\textsuperscript{a}Knowledge Exchange Division, The Sax Institute, Sydney, Australia; \textsuperscript{b}Australasian Cochrane Centre, Monash University, Melbourne, Australia; \textsuperscript{c}World Vision Australia, Melbourne, Australia; \textsuperscript{d}Evidence and Evaluation Branch, NSW Ministry of Health, Sydney, Australia

(Received 20 October 2013; accepted 15 January 2014)

The development of definitions is an integral part of the research process but is often poorly described. This paper details the iterative development of five definitions: Policy, Health policy-maker, Health policy agency, Policy documents, and Research findings. We describe the challenges of developing definitions in a large multidisciplinary team and the important methodological repercussions. We identify four factors that were most helpful in this process: (1) An emphasis on fit-for-purpose functionality, (2) Consultation with in-context experts, (3) Our willingness to amend terms as well as definitions, and to revisit some methods and goals as a consequence, and (4) Agreement that we would satisfy: accept ‘good enough’ solutions rather than struggle for optimality and consensus.

Keywords: definition; policy; policy-maker; knowledge exchange; research methods

What’s in a name? That which we call a rose by any other name would smell as sweet.

William Shakespeare, \textit{Romeo and Juliet} (c.1597)

In real life, unlike in Shakespeare, the sweetness of the rose depends upon the name it bears. Things are not only what they are. They are, in very important respects, what they seem to be.

Hubert H. Humphrey, Democratic politician and US Vice President (1966)

‘When I use a word’, Humpty Dumpty said in rather a scornful tone, ‘it means just what I choose it to mean – neither more nor less’.

Lewis Carroll, \textit{Through the Looking Glass} (1871)

Introduction

Ludwig Wittgenstein, the Austrian philosopher, illustrates the difficulties of defining terms in his exploration of the word \textit{game}. He argues that definitions of \textit{game} that...
focus on amusement or competition or rules each neglect import dimensions of the
many activities we call games, and thus no single definition can be found. But, he
suggests, we probably do not need a definition because we are sufficiently familiar
with enough things which are games and enough things which are not games to re-
ognize the difference between them (e.g. Canfield, 1981).

This is true for most of us most of the time. We don’t need to define concepts such
as friend, beauty, or irony1 because, like the famous US Supreme Court ruling on por-
nography (Jacobellis v. Ohio, 1964), we know it when we see it. We form views and
make decisions based on comparisons and approximations. We cannot know that oth-
ers perceive concepts in the same way, but we assume that in most circumstances it is
close enough. However, the use of terminology in research and evaluation requires a
more rigorous approach. Defining terms is one of the building blocks of research
design. It obliges researchers to be exact about the character and parameters of con-
cepts or phenomena under study, and to describe them transparently so that others
can assess their validity. It also facilitates effective communication within and across
fields of inquiry. This is particularly important in fields such as health research utiliza-
tion which have emerged from a range of research traditions across diverse disci-
plines (Rabin, Brownson, Haire-Joshu, Kreuter, & Weaver, 2008).

The development of a definition is an integral part of the research process. Con-
sequently, it is usually iterative and informed by the very process that it is intended
to support. Subjects of research become more complex and fragmented as they are
better understood, requiring revised distinctions and ever tighter definitions. This
increasing precision feeds back into the research process, provides a jumping off
point for further investigation, and can also be a valuable output of research. Yet
pecificity presents particular challenges, particularly in social research. Operational
definitions (constructs that define the tangible variables used as indicators in quanti-
tative research) strain to describe socially situated concepts or phenomena. Measur-
ing multidimensional constructs such as disability, anxiety, ethnicity, or intelligence
is hard; indeed, some argue that it is impossible (Altman, 2001; Bernard, 2000).
Such constructs defy universal application because they are embodied and perceived
differently by different people in different contexts. Conceptual definitions (brief
descriptions that tell us what a concept means) are more amenable. They refine con-
structs but can be formulated to cope with complexity and ambiguity, and they can
incorporate situational dimensions. But their very precision frequently limits their
utility to the conceptual lens and context of the study in which they were developed
making genuinely universal definitions something of a holy grail. For example,
Locock, Dopson, Chambers, and Gabbay (2001) found that opinion leaders can be
important mediators of research-informed organizational change initiatives, but
lament the limitations of researching this phenomenon more broadly due to the
improbability of capturing the multidimensional and contextual nature of an opinion
leader in a single definition.

Definitions also ‘frame’ their subject, asserting how it should be seen. As Hubert
Humphrey points out in the quote above, the terminology we use to describe a thing
affects our perceptions, including how we attribute value. This has critical implica-
tions not only for research, but for ‘real world’ responses too. For example, Huber et al. (2011) note that the current WHO definition of health – a groundbreaking con-
tribution to global health made in 1948 – now contributes to the medicalization of
society. Laderchi, Saith, and Stewert (2003) and Green (2006) demonstrate that
definitions of poverty affect policy agenda-setting, moral attribution, and resource
allocation. Further, Hodges (2008) argues that standardized definitions in ecology can impede progress by preventing important questions from being asked.

Similar concerns have been raised about terminology in the field of research utilization. The term evidence-based policy has been criticized as naïvely ignoring sociopolitical context and the need for negotiated decision-making in a pluralist democracy (Greenhalgh & Russell, 2009). Terms that suggest one-way linearity such as knowledge transfer are increasingly supplanted by more fluid and participative terms like knowledge exchange (Graham et al., 2006). And, despite definitions of the term knowledge translation that include concepts of exchange and multidimensionality (see, for example, WHO 2005 & NIDDR 2005 in Oborn, Barrett, & Racko, 2010), Greenhalgh and Wieringa (2011) argue that the metaphor embodied in the term constrains further study in the field by misrepresenting the socially constructed nature of how health knowledge is produced and used. It seems that terminology and definitions can provide precise in-a-nutshell conceptual syntheses that support critical debate and rigorous investigation, but they can also frame phenomena unhelpfully. The imperative is to select the best-fit terms and to define them clearly in relation to their context and use.

Some of the more granular research utilization terms such as policy, policy-maker, and policy agency are used in a variety of theoretical, empirical, and commentary articles without defining them. The danger is that, rather like Humpty Dumpty, their meaning is known only to the people using them. For example, few articles that explore policy-makers’ use of research explain the professional roles their study encompasses: are these policy-makers government employees? Political appointees? Elected politicians? Ministerial advisors? A mixture? It is important that we know because the power, constraints, and requirements of these roles differ considerably and are likely to affect how research is used (Haynes et al., 2012). But concepts which are politically volatile and strongly subjective such as policy are particularly resistant to definition (Barrett & Hill, 1984; Kemm, 2001). As Smith notes, ‘the struggle to classify or define policy is itself at the center of political conflict’ (2002). Some authors offer procedural definitions, e.g.

A long term, continuously used, standing decision by which more specific proposals are judged for acceptability in terms of means to be employed, ends to be pursued and time frame in which these proposals will have to fit. (Blum in de Leeuw, 1989)

Some opt for a less tangible approach which takes account of the inaction often associated with policy, e.g. ‘Something that one group of actors wishes to see carried out by others’ (Barrett & Hill, 1984) and ‘Whatever governments choose to do or not to do’ (Dye, 1984 in Fischer, 2003). While others focus on policy’s diffuse and values-focused characteristics, e.g. ‘Policy … consists of a web of decisions and actions that allocate … values’ (Ham, 1993 in Kemm, 2001).

Definitions of terms such as policy-maker will always require some contextual specificity because of the different systems in which policy work is done. Walton and Macagno (2009) argue that definitions are not ‘What a thing absolutely is,’ but ‘What a thing is commonly considered to be, based on evidential considerations pro and contra.’ As such, they are works in progress, evolving to reflect our changing knowledge of the world and adapted for situation-specific use. Ultimately, the value of a definition lies in its functionality for particular purposes (Chan & Donovan, 2005): given our goals, how well does the definition capture what we consider this thing to be? In some cases, the goal is to develop universal definitions that traverse
disciplines and contexts, in others it is to clearly unpack and boundary terms so that they can be investigated.

In this paper, we will describe the evolution of five definitions that were developed for the purposes of an intervention study known as SPIRIT (Supporting Policy In health Research: an Intervention Trial). SPIRIT was designed to test strategies for helping government health agencies increase their use of research evidence in policy and program development. Broadly, the study comprises a multifaceted interactive program of tailored education and resource provision, six measurement periods, and a process evaluation. It is being rolled out to six health agencies in Sydney, Australia over two years using a stepped wedge design.

We needed to define five terms for the purposes of this study: Policy, Health policy-maker, Health policy agency, Policy documents, and Research findings. We intended to use these definitions to: identify bodies of work from which policy/program documents could be selected for review; establish eligibility for an agency-level sampling frame and for individual participation in outcome measures; provide eligibility criteria for selecting documents to be reviewed in interviews; and to provide a frame of reference for participants completing outcome measures.

We have three aims:

1. To illustrate the value of fit-for-purpose conceptual definitions. We do not propose that others will necessarily find our definitions useful – they were developed solely for the purposes of our study – but we do urge research utilization investigators to explain how they defined their terms in published articles.

2. To provide an account of the challenging process of developing definitions. We hope this will counter depictions of apparently smooth and linear research processes that dominate the literature (Adler & Clark, 2011).

3. To share what we learnt during this process. In particular, the major considerations required to define these concepts, the value of the process of developing definitions, and the factors that facilitated this process.

Methods and results

The definitions were developed to be fit-for-purpose, i.e. to serve specific needs within the study. The purpose of each definition is outlined in Table 1, and their development and functionality are explored in turn below.

The process in all cases was iterative: two members of the team developed initial definitions based on broad searches in academic and government literature, and online reference sources. These were adapted in response to our developing study aims and constraints. Draft versions were reviewed by colleagues working in policy and program development and were revised accordingly except where the feedback did not align with the study design. Definitions were then reviewed by the whole investigator team, revised according to their advice, and signed off.

Technically, that is what happened, but it did not happen smoothly. Despite being aware of the importance of definitions, we were slow to engage in detailed dialog about them. But as soon as we moved into action – testing draft instruments, constructing sample frames, building consultation networks, etc. – the need to be precise about who and what was being sought became pressing. At that point, we found that in some cases no definitions existed, and in other cases they were plentiful but...
Table 1. Definitions and definitional purposes of terms used in a knowledge exchange intervention study.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policy</td>
<td>A formal statement or action plan developed by a government agency or statutory body in response to an identified problem. This includes statewide or national legislation, policies, programs, directives, protocols, guidelines, and service models</td>
<td>Identification of bodies of work from which documents could be selected</td>
</tr>
<tr>
<td>Health policy-maker</td>
<td>Someone employed in a policy agency who drafts or writes health policy documents or develops health programs, or who makes or contributes significantly to policy decisions about health services, programs or resourcing</td>
<td>Eligibility for agency sampling frame</td>
</tr>
<tr>
<td>Health policy agency</td>
<td>(A body within) a state or federal government department, or a statutory authority, whose focus is to develop policy which has an impact on statewide or national services and programs intended to improve individual, family or community health</td>
<td>Eligibility for agency sampling frame</td>
</tr>
<tr>
<td>Policy documents</td>
<td>A review, report, discussion paper, draft or final policy, formal directive, program plan, strategic plan, ministerial brief, budget bid, service agreement, implementation plan, guideline or protocol with a focus on health service or program design, delivery, evaluation or resourcing</td>
<td>Eligibility for selection of documents for targeted interviews</td>
</tr>
<tr>
<td>Research findings</td>
<td>Analyses of quantitative or qualitative data, or theory, found in peer reviewed papers, technical monographs or books, or in grey literature such as internal studies and evaluations, and reports on authoritative websites</td>
<td>Inform the scoring of the outcome measures</td>
</tr>
</tbody>
</table>

divergent. Some of the richest sources were of uncertain quality, e.g. Wikipedia, blogs, online dictionaries. Then, as we started to consult with colleagues in policy and program work about the face validity of draft definitions, their feedback raised fundamental questions that caused us to critique and amend aspects of the intervention or evaluation design, so we bounced back and forth between the literature, the evolving study design and consultation advice. Our team comprised a large and diverse group of researchers and policy-makers located across multiple institutions, cities and countries making it difficult to find time and the appropriate forum to debate disagreements and find a resolution (see also Billings' [2004] account of ‘angst-ridden’ definition development processes within a large research team). We recognized this as an essential and familiar developmental process, but there were differences of opinion about its duration: at what point would we draw a line in the
sand? In fact, we never reached the fabled end point. Definitions that had been ‘signed off’ continued to change in response to shifting perceptions of the study design, goals, contexts, and participants right up to the moment we were obliged to give them to our participants. Some are still contested, as we describe below.

**Policy**

A formal statement or action plan developed by a government agency or statutory body in response to an identified problem. This includes state-wide or national legislation, policies, programs, directives, protocols, guidelines, and service models.

Our primary purpose in defining *policy* was: (a) to identify streams of work from which policy documents could be identified and (b) to include the many types of policy and program work conducted by participant agencies because each of them had the potential to influence population health and health services in important ways.

We arrived at our final definition through a process of plunder. We reviewed definitions of *policy* found in dictionaries and in the academic literature, chopped them up and rearranged them to synthesize aspects that aligned with our aims, and dropped or amended aspects that didn’t. For example, consultations with some colleagues who were developing important population level initiatives indicated that they did not regard their work as a form of policy, so we supplemented the definition with examples to capture these: programs, protocols, guidelines, and service models. Given that our intention was to identify policy documents, the definition purposefully focused on formal statements, or action plans that would have written materials associated with them. The definition does not demand that policy includes action because our intervention focused on getting research into policy and program development processes, but did not extend to implementation or health services practice.

The research team used this definition behind the scenes, often as a way of thinking about health policy agencies because these agencies were defined, in part, by their function and tasks. However, the definition appeared to have minimal utility for study participants. It was included with some study information, especially in relation to the selection of policy documents for review in structured interviews, but a definition was neither asked for nor referred to in any interviews or clarifying conversations. Presumably, policy-makers have a better idea than we do of what constitutes policy and do not require a definition. They know it when they see (or do) it.

**Health policy-maker**

Someone employed in a policy agency who drafts or writes health policy documents or develops health programs, or who makes or contributes significantly to policy decisions about health services, programs or resourcing.

This definition was used to identify individuals within participating agencies. Specifically, to differentiate between staff who would be eligible or ineligible to take part in the outcome measures. Our primary goal was to ensure the intervention reached people who were in the best position to benefit from it – i.e. who had sufficient opportunity to apply research-related knowledge, skills, and resources in their work – so this was where we intended to look for measureable change.

Dictionary definitions of *policy-maker* lacked the specificity we required, e.g. ‘A person responsible for or involved in formulating policies, especially in politics’
Oxford Dictionaries) or ‘Someone who sets the plan pursued by a government or business etc.’ (The Free Dictionary). The academic literature did not present the required detail either since it tends to define the term by who it includes and excludes rather than describing it (e.g. DeRoeck, 2004; Feldman, Nadash, & Gursen, 2001; Haynes et al., 2012).

Early consultation with policy colleagues alerted us to the breadth of roles involved in what we were calling policy-making. They advised us to include staff at different levels within each agency, and to ensure that staff who focused on population-level program development and resourcing were included. This advice, together with our observation that the people we were calling policy-makers did not use that term themselves, caused us to reconsider the term policy-maker. We liked Maureen Dobbins and colleagues’ term decision-maker (e.g. Dobbins, Jack, Thomas, & Kothari, 2007) for its inclusivity, but recognized that this was advisedly broad so as to include health services managers – a group who were excluded from our study. We were also mindful of Weiss’s (1982) finding that few policy-makers are in the position to make final policy decisions. This is because policy development is a collective incremental process characterized by mutual adjustment, move and counter move, accretion and negotiation, thus policy-makers focus on affecting the shape and content of policy discourse rather than deciding on overarching policy. Consequently, given that the term was to be employed primarily for internal use (when we used the definition in agencies it appeared under the heading of eligible participants), and the term policy-maker was dominant in the literature, and we continued to use it.

The consultation also raised the murky question of what policy-makers actually do – what component tasks make up the policy-making process and how could we identify those that are most amenable to research input? This included tasks such as refining/analyzing policy questions, reviewing data, conducting or commissioning research or evaluations, consultation with internal and external stakeholders, working with advisory committees, and information synthesis. The catchall phrase … contributes significantly to policy decisions … was added to capture that breadth of activity.

We found this to be a functional definition for our purposes, and feedback from participating agencies confirmed that it provided clear guidance for assessing staff eligibility. For example, it was precise enough to exclude contractors from the study and staff who did not contribute to policy or program development such as those in administration and operations, but inclusive enough to capture many strata of policy and program developers and to allow pilot testing in an agency that funds state health services.

**Health policy agency**

(A body within) a state or federal government department, or a statutory authority, that focuses on developing policies or programs designed to improve state-wide or population level health.

We needed to define what a health policy agency was in order to develop a sample frame for organizations that were to be invited to participate in the intervention. Rather like Wittgenstein’s game example above, in the early design stages we knew enough organizations that were like our conception of a health policy agency and enough that were not like it to categorize organizations as potential participants or not, but we were unable to find any definitions in the literature from which to draw
rigorous criteria that could be used to create a sample frame. Consequently, we
developed a pragmatic definition that matched study eligibility requirements with
the goals and in-development methods of the intervention. Our constraints (geo-
graphic and financial) were front of mind in this process.

It was agreed that a health policy agency must develop population level health
policy or programs as its core business. This excluded health agencies whose pri-
mary role is operational, e.g. accreditation, compliance, and benefits schemes admin-
istration. Given our intervention focused on public health and clinical research, we
took a narrow perspective of health which excluded many health-related agencies
such as community services. Non-government organizations could have been
included in this definition; however, in our study we wanted to test the intervention
with larger agencies that had most scope to directly influence population level health
– so only state or national level government organizations and bodies with statutory
authority were considered.

This definition worked well enough for our purposes, particularly after discus-
sion with agency staff in the early phase of the study led us to refine it to give
greater weight to program development, and to remove some unnecessarily granular
description.

Policy documents

A review, report, discussion paper, draft or final policy, formal directive, program plan,
strategic plan, ministerial brief, budget bid, service agreement, implementation plan,
guideline or protocol with a focus on health service or program design, delivery, evalu-
ation or resourcing.

The purpose of defining the term policy document was to provide inclusion criteria
that would enable each participating agency to nominate four eligible documents
which best represented their use of research evidence in policy or program agenda-
setting, development, implementation, or evaluation within a given time period. Key
personnel were then interviewed about the process of developing these documents
which were considered to be proxies for the policies and program that they related
to.

The definition had to be broad enough to allow agencies to select documents that
reflected their work, but contained enough to identify the sorts of documents that
could usefully be informed by research. The consultation process involved col-
leagues in policy and program development kindly sending us examples of their
documents as well as commenting on the draft definition. With each consultation
our overarching definition diminished and our list of examples became longer until,
eventually, only a list was left. The range and diversity of authors, content, aims,
and intended readers were so great that we were unable to describe these dimensions
meaningfully for all eligible agencies in a single brief definition. Happily, we were
told that concrete examples were preferable so we honed these and, following a fur-
ther round of consultation, we dropped press releases from the list (they lacked
detail and were likely to be less ‘authentic’), and added budget bids.

In practice, as in development, this was our most problematic definition. From
the start, agencies repeatedly required verbal clarification and expressed confusion.
No specific problems with the definition itself were identified, and so no improve-
ments were made, but the concept of a policy document remained ambiguous with a
need for locally specific discussion in each site.
Research findings

Analyses of quantitative or qualitative data, or theory, found in peer reviewed papers, technical monographs or books, or in grey literature such as internal studies and evaluations, and reports on authoritative websites.

The purpose of this definition was to inform the design of the outcome measures (i.e. to identify what sort of research-related information was being used at each measurement point) and to provide a common reference for the SPIRIT team and participants, particularly during interviews that focused on research use.

Definitions of research abound, but they tend to focus on the process of conducting research (systematic gathering and analysis of information to advance understanding) rather than the outputs of the process. Discussions of what constitutes research data provided a useful starting point (e.g. Australian National Data Service, 2011) as they drew attention to the important difference between raw and analyzed data. Given that most policy-makers do not have research qualifications (although a surprising number do), the intervention encouraged the use of analyzed data only, and the definition reflected this.

The goals of the intervention were to increase the use of research, but we took a pragmatic view of what that might entail. Much so-called gold standard research is neither relevant nor applicable to policy and program development, or cannot be produced within the necessary timeframes, and grey literature (e.g. government white papers, internal program evaluations) is often particularly helpful. Therefore, we sought a definition that was broad enough to encompass this. The definition did not need to address research quality because that was explored conversationally during interviews conducted as part of the study’s outcome measures.

We considered including advice from researchers in this definition, but this was rejected because it was too amorphous a construct and hard to qualify without talking to the researchers themselves. If policy-makers sought or used advice from researchers, this was captured in our survey forms and interviews. The definition provided in those instruments includes the addendum: Advice from researchers is considered to be research-informed information, but not research per se.

Few participants asked how we were defining research or referred to it during interviews, but several mentioned their surprise at the inclusivity of the definition given that we were researchers (and thus, by implication, attached to academic hierarchies of evidence). So we concluded that the definition was helpful primarily as an invitation for policy and program developers to talk about the ‘real world’ range of research they used.

The term itself – research findings – was contested. Some members of the team wanted to use the term research as a catch-all for the process of conducting research and for its outputs – the context made it clear which was which. But others argued for greater precision. We also debated using the term evidence. This debate was more challenging since it obliged us to explore our beliefs about the nature and status of human inquiry. Some argued that the term evidence is rhetoric (see Greenhalgh & Russell, 2009; Russell, Greenhalgh, Byrne, & McDonnell, 2008). Others argued that it was common parlance and that authors such as Head (2008, 2010) and Klein (2003) had convincingly expanded the term to include other policy-relevant forms of information, but for the sake of clarity we should use the term research evidence or evidence from research. We agreed on the term research findings for use in this study but note that it was used inconsistently (as this very paper illustrates).
Discussion

We managed to arrive at five definitions or example lists that were largely successful for our specific purposes which were to: identify bodies of work from which policy/program documents could be selected for review; establish eligibility for a sampling frame and for participation in outcome measures; provide eligibility criteria for selecting documents for review during interviews; and to provide a frame of reference for participants completing outcome measures. The definitions also played an important role in the design of the intervention and outcome measures. Being explicit about the concepts we were studying obliged us to confront issues that had remained ambiguous in our day-to-day deliberations.

Our draft definition of policy was a pragmatic amalgam of dictionary definitions. Following consultation, we supplemented it with examples of more specific statewide population health and clinical initiatives. With the benefit of hindsight, a more inclusive response would have been to position program development work more explicitly within the term itself so that we were referring to policy and program development rather than policy. This would also have been a more process-orientated perspective which could have helped us to focus on the range of work practices we wished to influence rather than on outputs which we were already targeting with our definition of policy documents. A similar argument can be applied to our other terms. Each definition expanded to include program development, but the terms themselves did not. For example, having been advised by potential participants who we were calling policy-makers that they did not refer to themselves that way, we wrestled with alternatives to the term. We were unable to agree on a single term but, in the light of subsequent changes, policy/program developer may have been more suitable for our purposes. Likewise, the definition of health policy agency worked well, but if we were to use it with participating agencies we would consider revising the term to reflect the importance of program development: health policy/program development agency.

Our definition of policy documents (which we would now call policy or program documents) was designed to help participants identify key documents produced in the course of policy/program work that would be discussed in interviews. Variation between agencies in work processes and document types precluded a description that could be applied to all; therefore, the definition became a broad list of examples. However, even this list remained problematic since there was no standard nomenclature: one agency’s discussion paper is another agency’s scoping brief. It is no surprise that, despite our changes to the instructions for applying this definition and the additional inclusion criteria, agencies continued to struggle with it. We recognized that a standalone definition that would apply to all agencies was probably not possible. Given that the process of presenting documents for assessment was an unfamiliar and potentially uncomfortable form of participation, we had an obligation to talk with the people nominating documents about how to apply the definition in their context, and to situate this within a broader conversation about our methods and aims and how the data would be used. Agencies had been given this information in written form, but more nuanced dialog that took account of local factors and concerns was also important.

Some may disagree with our decision to include grey literature in the definition of research findings, but this reflected the real world use of investigative information within policy, and was acknowledged as such by participants. Including grey literature
resulted in in-depth conversations about its use which, in turn, educated the research team about the variety and quality of innovative program designs and fit-for-purpose evaluations conducted by many government health agencies, most of which are never formally published. It was a reminder that much grey literature is grey not because of its merit, but because it is sometimes generated using expeditious methods, or is developed in contexts that do not prioritize academic publishing.

The process of developing definitions – reviewing the literature and other likely sources, synthesizing and adapting definitions through the lens of the study methodology and goals, and consultation (testing) with colleagues in policy/program development – was messy but productive. We identified four factors that contributed to this productivity:

1. An emphasis on functionality: our definitions were intended to be purpose specific so we were able to focus exclusively on our study aims rather than striving for generalizability. We believe that our definitions of policy, health policy-makers, and health policy agencies are broadly applicable in different contexts, but they should be reviewed for functionality when used in other studies.

2. Input from a broad range of people with in-context expertise. This helped us to understand the phenomena we were exploring, to refine the parameters of our inquiry and articulate the definitions so they reflected this. The appreciation that program development needed to be recognized explicitly within the field that we were calling policy-making led to changes in the definitions and, hopefully, to more sensitive communication with agencies who were participating in the intervention.

3. We asserted that conceptual definitions can and should be refined in response to data collection. Frustrating as it sometimes felt, the iterative revisions to the definitions obliged us to critique the terms we were attempting to define and, in some cases, to amend particular study aims and methods. We are confident that this resulted in more rigorous definitions, terms and methods. We would have preferred a clearer end-point – greater certainty about having arrived at the ‘right’ terms and definitions – but this will always be something of a judgment call since each application is liable to test them in new ways. Wading around in the marshlands of conceptual thinking and applied problem-solving is in the nature of social research and is critical for sound methodological development, but it must be balanced by the pragmatic need to arrive at an agreed (if imperfect) final decision. We probably underestimated how lengthy this process would be given the complexity of our study, its uniqueness, and the contexts in which it was to be applied. With this in mind, earlier consultation about our definitions and use of a more systematic and transparent revision and consultation process would have been valuable.

4. We agreed early on that the chance of reaching consensus on each of the definitions was, to quote one of the investigators, roughly that of a snowball in hell. The definitions would be intersubjective rather than objective, that is, they would reflect a variety of broadly agreed (or at least explicitly recognized) perspectives rather than providing a factual description (Gillespie & Cornish, 2010). The concept of deliberative dialogue was used where the goal is to explore different perspectives and to reach a shared understanding.
of the topic that can feed into solutions (Boyko, Lavis, Abelson, Dobbins, & Carter, 2012). This allowed us to move forward with some terms and definitions that were not favored by every team member. In fact, the debate about their merits and implications continues.

What is in a name? Returning to our opening quotes, we would argue that Hubert’s advice is more persuasive than Romeo’s declaration. The name that we give a thing frames how we see it. Defining a thing makes this frame more transparent and, certainly in our case, can lead to helpful critical consideration of the name itself. Although it is valuable, unpacking and delineating complex social phenomena is a challenging task, even within the confines of a relatively focused study. We take our hats off to those who tackle the fraught goal of developing universal definitions.

Conclusion
This paper explores the development of fit-for-purpose conceptual definitions of five terms: policy, health policy-maker, health policy agency, policy documents, and research findings. It illustrates the role of definitions as one of the building blocks in study design and provides an account of the process of developing them – reviewing the literature and other likely sources, synthesizing and adapting definitions through the lens of the study methodology and goals, and testing draft definitions with colleagues in policy and program development. It also describes our learning; in particular, our increasing appreciation of the value of the process itself, and the four factors that were most helpful: (1) An emphasis on fit-for-purpose functionality, (2) Consultation with in-context experts, (3) Our willingness to amend terms and definitions during their early stages of use and to refine some methods and goals as a consequence, and (4) Agreement that the team would ‘satisfice’: accept ‘good enough’ solutions rather than struggle for optimality and consensus.

The development of definitions was a challenging, messy, and invaluable process. It allowed us to communicate more effectively with study participants, but it also obliged us to recognize our hazy understanding of key concepts and to hone and articulate them, to test our assumptions about shared meanings, and to wrestle with some fundamental questions about the nature of our inquiry. All of which, we believe, improved the overall rigor of the study. We ask others in the field of research utilization to define their terms when publishing.

Funding
This work was funded as part of the Centre for Informing Policy in Health with Evidence from Research (CIPHER) an Australian National Health and Medical Research Council Centre for Research Excellence [grant number APP1001436] and administered by the University of Western Sydney.

Notes
1. Some members of the research team argued that this point is poorly made given the (ironic) success of Alanis Morisette’s 1996 hit single in which she erroneously claims that events such as rain on one’s wedding day, dying soon after winning the lottery, and failure to take good advice are ironic. They are, of course, merely unfortunate.
2. We did identify problems with the instructions and additional criteria we gave about how to select the documents, but this was not an aspect of the definition and so is not discussed here.

Notes on contributors
Abby Haynes is a research officer at the Sax Institute in Sydney, and a research affiliate at the School of Public Health, University of Sydney. She is currently investigating the relationship between research and policy-making, striving to understand how interventions in complex systems can be improved, and pursuing her work using a variety of qualitative research and evaluation methods.

Tari Turner is a senior health advisor at World Vision Australia and an adjunct senior research fellow at Monash University. She is passionate about translating knowledge into practice and, in particular, making sure women, babies, and children across the globe benefit from the findings of health research.

Sally Redman is the CEO of the Sax Institute in Sydney. She is a public health behavioral scientist and her primary interest is in increasing the use of research evidence in health policy and practice. The Sax Institute seeks to assist health decision-makers to make better use of findings from research.

Andrew Milat is an associate director of Evidence and Evaluation Branch in the NSW Ministry of Health, Australia. He is a policy-maker-come-intervention researcher with over 16 years’ experience in the design, implementation, and evaluation of innovative health policies and programs at national, state, and local levels. He has overseen the development and implementation of numerous public health research programs and intervention trials, and is a strong advocate for the application of research evidence into policy and program development.

Gabriel Moore is the principal analyst in the Knowledge Exchange Division of the Sax Institute. She is currently completing a PhD on the effectiveness of knowledge brokering in health policy settings. She is the lead author of What works to increase the use of research in population health policy and programmes: a review (2011).

References


Appendix 10. Chapter Seven Manuscript

Figuring out fidelity: a worked example of the methods used to identify, critique and revise the essential elements of a contextualised intervention in health policy agencies
Background: In this paper, we identify and respond to the fidelity assessment challenges posed by novel contextualised interventions (i.e. interventions that are informed by composite social and psychological theories and which incorporate standardised and flexible components in order to maximise effectiveness in complex settings). We (a) describe the difficulties of, and propose a method for, identifying the essential elements of a contextualised intervention; (b) provide a worked example of an approach for critiquing the validity of putative essential elements; and (c) demonstrate how essential elements can be refined during a trial without compromising the fidelity assessment. We used an exploratory test-and-refine process, drawing on empirical evidence from the process evaluation of Supporting Policy In health with Research: an Intervention Trial (SPIRIT). Mixed methods data was triangulated to identify, critique and revise how the intervention’s essential elements should be articulated and scored.

Results: Over 50 provisional elements were refined to a final list of 20 and the scoring rationalised. Six (often overlapping) challenges to the validity of the essential elements were identified. They were (1) redundant—the element was not essential; (2) poorly articulated—unclear, too specific or not specific enough; (3) infeasible—it was not possible to implement the essential element as intended; (4) ineffective—the element did not effectively deliver the change principles; (5) paradoxical—counteracting vital goals or change principles; or (6) absent or suboptimal—additional or more effective ways of operationalising the theory were identified. We also identified potentially valuable ‘prohibited’ elements that could be used to help reduce threats to validity.

Conclusions: We devised a method for critiquing the construct validity of our intervention’s essential elements and modifying how they were articulated and measured, while simultaneously using them as fidelity indicators. This process could be used or adapted for other contextualised interventions, taking evaluators closer to making theoretically and contextually sensitive decisions upon which to base fidelity assessments.

Keywords: Fidelity, Essential elements, Flexibility, Process evaluation, Intervention theory
Background

The process evaluation literature frequently characterises interventions as a ‘black box’ meaning that little is known about how they function, including the hypotheses that underpin their design [1–3]. Process evaluation shines a light in this box by investigating ‘how and why’ questions about the intervention’s implementation, change mechanisms and contextual interactions [4].

Fidelity assessment is a fundamental part of process evaluation. Its purpose is to ascertain ‘the degree to which an intervention or procedure is delivered as intended’ ([5]: 407). This is achieved by operationalising the intervention theory and monitoring the consistency and congruence with which it is implemented [6–9]. In order to determine if the delivery was ‘as intended’ two areas of assessment should be considered: implementation fidelity and theoretical fidelity. Implementation fidelity tells us to what extent the intervention-as-delivered matched the intervention-as-planned. The assessment focuses on measurable or codifiable dimensions such as how intervention providers were recruited and trained, what proportions of targeted people were reached, the amount of exposure participants had to intervention activities (intervention intensity) and the consistency with which the intervention components were delivered in each setting [10]. This is a comparative enquiry that identifies variation between desired and actual activities, between participant sites and over the duration of the intervention. Implementation fidelity assessment is vital for understanding the intervention’s variation [9, 11], determining its feasibility [6, 12] and determining whether an ineffective intervention was due to poor implementation or flawed design [3, 12–15].

Theoretical fidelity tells us the extent to which the intervention-as-delivered was congruent with the intervention theory (the logic and hypotheses that underpin the intervention design [16–18]). This intervention theory is operationalised in the form of ‘essential elements’: manifestations of the theory—the ‘active ingredients’—which must be implemented if the intervention is to be effective [2, 6]. The assessment uses the intervention’s essential elements as indicators for a formative enquiry that makes judgements about the validity of the intervention design in practice. This helps us determine how the intervention worked or why it did not [17–19]. As the new UK Medical Research Council guidance for process evaluation states,

It may never be possible to fully understand how variations in delivery affect outcomes, given that adaptations do not occur at random, and will be confounded by factors promoting or inhibiting intervention effects. A strong understanding of the theory of the intervention is a prerequisite for meaningful assessment of implementation, focused not just on the mechanics of delivery, but whether [the] intervention remained consistent with its underlying theory ([4]: 41).

Ensuring theoretical fidelity is vital for assessing the program theory [14], predicting outcomes [9, 20, 21], translating and adapting interventions for other contexts [12, 19, 22], further developing the intervention’s evidence base [9, 23] and enabling ‘streamlining’ that may reduce burden and cost [6, 24]. In trials of complex interventions, fidelity assessment supports interpretation of intervention outcomes ensuring that observed effects (or lack thereof) can be linked to implementation of the intervention. More positive outcomes have been observed when interventions are delivered with high implementation and theoretical fidelity [9, 12, 18], including in flexible interventions providing that adaptations are locally and culturally appropriate and are congruent with the program theory [11, 25–28].

The concept of assessing fidelity as part of intervention evaluation originates from psychotherapeutic programs. The aim of fidelity assessment in this context is to ensure prescribed treatments are delivered with minimal variation [15, 21] and adhere to the behaviour-change theory that informed their design. This approach has proliferated within implementation science and is now used for a range of interventions designed to change professional practice in health care. There is increasing formalisation of the theory that underpins these interventions and their essential elements, leading to testable theoretical frameworks and taxonomies of standardised techniques that support replicability and evidence synthesis across studies, e.g. [29, 30].

However, this approach cannot be used for all intervention trials. Indeed, its proponents do not suggest that methods designed to assess the fidelity of clinical actions performed by healthcare workers in the process of delivering healthcare [30] should necessarily be more widely applied [31]. Two aspects in particular pose problems for translation: (i) the focus on individual behavioural change and (ii) the specificity with which the theory is operationalised. The former is problematic because the best-developed methods of fidelity assessment identify essential elements from a taxonomy of techniques derived from individual behaviour-change theory [29, 32]. No equivalent exists for interventions informed by broader social science theories that target complex interactive, organisational and system level properties [10, 33, 34]. The latter is problematic because it is too restrictive for assessing the fidelity of flexible interventions designed to allow local adaptation in order to increase their relevance and applicability [35–37]. Nor does it capture how interventions respond reflexively to
unique characteristics and unpredictable reactions in their settings [38]. This fidelity/adaptation dilemma [22] is particularly pertinent for interventions based on composite theory that are designed for dynamic real world systems in which it is necessary to balance standardisation of both form and content with responsiveness to context. Indeed, resolving the fidelity/adaptation dilemma in these contextualised interventions is one of the most important challenges for evaluation [39]. (For clarity, we use the term contextualised intervention rather than complex intervention in this paper as complex interventions are most commonly defined in relation to structural design rather than their theoretical or contextual characteristics [40].)

A growing body of literature documenting the evaluation of contextualised large-scale interventions attempts to tackle the challenges of composite theory, flexibility and responsiveness to context. These interventions include those informed by ecological, complexity, empowerment and realist perspectives, and those tailored by local providers or developed participatively, e.g. [35, 41–48]. However, while many studies link their intervention’s essential elements to theory, they seldom report sufficient detail for others to see how that theory was translated into specific intervention techniques (rather than other techniques or variants that might be equally well supported by the theory). Moreover, some assume prior knowledge of the form that the intervention and its underlying theory will ultimately take, failing to acknowledge that an intervention’s so-called essential elements may function as conditional elements: contingent on the interaction between intervention techniques, heterogeneous participants and contextual characteristics [49–52]. Consequently, the intervention designers may be obliged to make countless incremental adjustments to the techniques and the theory that underpins them while the trial is in progress; thus, ‘By the end of the program, the designers’ operating theory may look quite different from the theory with which they started’ [53]. Intervention studies targeted at community populations such as cultural groups often highlight the contingent validity of program theory and why it should be critiqued, (re)operationalised and potentially rejected, depending on local needs and conditions, e.g. [27, 48, 54], but this is often lacking in organisational level studies [51]. So few trials conducted in policy organisations have been reported that, currently, our knowledge of how intervention strategies may interact with variations in these environments is little more than speculative.

Despite widespread agreement that all intervention trials should document the extent to which their essential elements were delivered [6, 12, 36], no universal methodology exists for identifying or measuring essential elements [8–10, 55] and, for interventions with composite theory, there is sparse guidance for ensuring putative essential elements are valid indicators of the underpinning theory [9, 20, 38, 55]. So how should we determine which elements of an intervention are genuinely essential and which can be adapted without impairing effectiveness? Calls for greater attention to these questions are widespread, coming from multiple sectors in health [5, 6, 13, 17, 23, 38, 56, 57], education [19, 55, 58] and community development [11, 20, 35, 59].

How are essential elements identified?

When based on previous studies, intervention designers can identify essential elements from analysis of earlier interventions or operationalise them using exemplary models that have established effectiveness [9, 10, 12]. Theoretically informed standardised behaviour-change techniques are in development, but these are currently limited to interventions founded on psychological theories [30]. When designing and evaluating novel contextualised interventions, designers can either articulate the essential elements themselves or consult with expert colleagues [8, 9, 19, 56]. Many evaluations tackle this post hoc, piecing together the essential elements via discussion with the designers and/or by reviewing intervention materials [12, 19, 55].

The design of interventions in trials is often founded on an amalgam of hypotheses that attempt to take account of inter-related theoretical, contextual and pragmatic factors. These include formal and substantive theories; hunches based on professional experience; and considerations such as study resources, demands on participants, existing practice and infrastructure constraints. The intervention’s essential elements are representations of these composite working hypotheses [55]. Thus, essential elements are not extant change agents waiting to be discovered; rather, they are ways of putting working theories into practice in particular circumstances, chosen as the ‘best bet’ from many potential candidates [7]. It is not surprising, therefore, that newly developed essential elements for all types of intervention need to be assessed in situ to determine the extent to which they capture and truly deliver the intervention theory in the context of messy real world delivery [17].

How specific should essential elements be?

The degree to which essential elements are specified must align with the level of flexibility in the intervention design. Minimally specified essential elements are appropriate for highly flexible interventions because they can be interpreted for different contexts [34, 60, 61]. These essential elements tend to be expressed as principles, goals or functions (rather than specific techniques or formats) as these provide scope for diverse implementation strategies. Fidelity rests on the extent to which the
resulting strategies align with the principles, goals and/or functions (see [59] for examples) [33, 62]. Equal emphasis should be placed on how discretionary elements were tailored and with what process effects [33, 59].

Where the intervention combines standardised and flexible components, an appropriate balance must be found. Essential elements that are too tightly specified oblige providers to adhere to prescriptive scripts and techniques which may be suboptimal or entirely inappropriate in different contexts and circumstances [27, 35, 62], whereas minimally specified essential elements may not provide sufficient concrete guidance for developing or monitoring the core intervention activities [21]. The specificity of essential elements is critical for defining what the intervention is and what it is not, including which elements are genuinely essential and which can be adapted [13, 55]. To date, the literature does not provide the detail needed to identify, or determine the specificity of, essential elements for contextualised interventions.

**Aims**

In this paper, we identify and respond to the challenges of fidelity assessment in contextualised interventions using the Supporting Policy In health with Research: an Intervention Trial (SPIRIT) study as an example. SPIRIT is testing the effects of a suite of strategies designed to increase the capacity of health policy agencies to use research. SPIRIT recognises that policymaking is a messy subjective social process that takes place in complex open systems with myriad influences [63]. How research is used in policymaking is not fully understood [64], but it appears that different structures, pressures, relationships, values and events interact to shape its relevance, applicability and use, and that this flux cannot be controlled during interventions [22, 43, 64, 65]. Consequently, SPIRIT draws on diverse theories from social and political science, targets individual and system level capacities and, as Table 1 shows, attempts to balance standardisation with responsivity to context in its implementation and evaluation.

Specifically we (a) describe the challenges of, and propose a method for, identifying the essential elements of a contextualised intervention (a semi-flexible, theoretically eclectic intervention designed for complex settings); (b) provide a worked example of an approach for critiquing the validity of putative essential elements; and (c) demonstrate how essential elements can be refined during a trial without compromising the fidelity assessment. We consider how this approach might complement current methods for identifying essential elements.

**Context for this study: SPIRIT**

Our fidelity assessment was developed and conducted as part of the process evaluation of Supporting Policy In health with Research: an Intervention Trial (SPIRIT). In this trial, six health policy and program agencies based in Sydney, Australia, participated in an intervention designed to increase the capacity of policymakers and program developers to use research in their work. SPIRIT was informed by cognitive behavioural theory, systems thinking, the literature on research utilisation, organisational change and adult learning theories. These were articulated in the form of the SPIRIT action framework (Fig. 1) and a list of change principles (Table 2) which, in turn, guided the intervention design and the goals and strategies of individual activities [63, 66].

The intervention comprised multiple components: (i) audit, feedback and goal setting; (ii) a leadership program; (iii) organisational support tools; (iv) the opportunity to test systems for accessing research; (v) research access; and (vi) educational symposia. These components had varying degrees of flexibility as outlined in Table 1. Agency staff received approximately 11 face-to-face sessions over the 12-month intervention period, combined with periodic feedback and ongoing access to resources. Proximal and distal outcomes included (1) organisational capacity to use research (staff knowledge, skills and perceptions of the value of research and organisational support for the use of research as demonstrated through leadership support, policies, tools and systems), (2) research engagement (accessing, appraising and generating research, and interacting with researchers), and (3) research use in policy or program work (demonstrated through the assessment of nominated policy documents). Agencies could prioritise outcomes they wished to improve by tailoring the intervention, e.g. to target particular knowledge or skills.

High-profile policy and research experts were recruited to deliver the face-to-face intervention sessions. The outcome measures comprised an online survey and two structured interviews. Further details are provided in the study protocol [66].

**The challenges**

Several characteristics of SPIRIT presented challenges for fidelity assessment. Addressing these challenges drove the methods we used:

1. **Composite theory.** The intervention was built on cross-disciplinary composite theory that had not been operationalised in previous trials. This theory was articulated in the SPIRIT action framework and change principles (Fig. 1 and Table 2), but these did not identify which intervention elements should be used as fidelity indicators, nor the level of specificity with which they should be operationalised.
The manner in which the essential elements should be articulated was complicated by the paradigmatic tensions and different fidelity traditions in the composite theory. For example, cognitive behavioural theories lean towards positivism and experimental intervention approaches and fall within the standardised approach to fidelity assessment outlined at the beginning of this paper in which essential elements are tightly specified. Systems thinking, on the other hand, proposes a complexity-orientated ecological worldview in which interventions are loosely specified for local adaptation and essential elements are articulated as principles rather than concrete techniques. SPIRIT, like many contemporary interventions, was occupying a middle ground.

2. Flexibility. The expression of the essential elements needed to accommodate three levels of flexibility: (a) agencies were able to select different session options from a menu of components, (b) they could tailor the topics and goals of these options to address local priorities, and (c) expert providers determined the detail of delivery (see Table 1). We could not foresee how these decisions would shape the content and form of the intervention. Given that meaningful comparison of the extent to which essential elements were delivered required that they be equally applicable across all intervention

| Table 1 The degree of flexibility in SPIRIT intervention components and subcomponents |
|---------------------------------|-------------------------------------------------|-----------------------------|
| Intervention components (fixed) | Subcomponents                                    | Targeted participants       | Degree of flexibility in form and content* |
| 1. Audit, feedback and goal setting | a. Feedback forum                          | Senior leaders and other key managers, as determined by each agency | Partial: Tailored presentation based on agency’s audit data. Informal discussion shaped by participants’ interests. |
|                                  | b. Intervention selection                    |                            |                                          |
|                                  | c. Identification of other strategies        |                            |                                          |
|                                  | d. Mid-intervention feedback                 |                            |                                          |
|                                  | e. SPIRIT newsletter                         | All agency staff involved in policy/program work | Partial: Tailored to each agency based on their audit data |
| 2. Leadership program            | a. Supporting organisational use of evidence | Senior leaders and other managers depending on size and configuration of agency | Partial: Standardised presentation determined in consultation with providers, but with agency-specific case examples. Discussion shaped by participants’ interests. |
|                                  | b. Leading organisational change             |                            |                                          |
| 3. Organisational support for research | a. Quarterly email endorsement of SPIRIT from agency’s CEO | All agency staff involved in policy/program work | Partial: Proforma text that CEOs can adapt |
|                                  | b. Access to WebCIPHER (an interactive research portal) |                            |                                          |
|                                  | c. Resources for improving the agency’s use of research |                            |                                          |
| 4. Opportunity to test systems for accessing research and reviews (brokered services) | a. Brokered commission of: a rapid systematic review OR an evaluation plan OR an analysis of linked data | Primary: Agency-selected staff who would benefit from experience commissioning a service. Secondary: all staff working in the topic area addressed by the commissioned product | Extensive: Standard brokerage processes are used but agencies choose the product, and specify the topic and how it should be approached to best meet their needs. |
|                                  | b. Summary of systematic reviews             | All policy/program staff working in the topic area covered by the review | Partial: Agencies choose the topic. They can nominate a particular provider and negotiate the form of the session. Providers shape the delivery details. |
| 5. Research access               | Three occasions of research access from two modes: | All policy/program staff working in the topic area covered by the forum | Extensive: Agencies choose the topic. They can nominate a particular provider and negotiate the form of the session. Providers shape the delivery details. |
|                                  | a. Interactive forums with researchers AND/OR |                            |                                          |
|                                  | b. Summary of systematic reviews             | All policy/program staff working in the topic area covered by the review | Partial: Agencies choose the topic |
| 6. Educational symposia for staff | a. Valuing research symposium                | All agency staff involved in policy/program work | Limited: Agencies can nominate case examples |
|                                  | b. Two symposia from: Access to research Appraising research Evaluation Working with researchers | All policy/program staff who might benefit from enhanced skills in the techniques covered | Partial: Agencies select symposia topics from the menu. They can tailor the focus, nominate case examples and providers, and negotiate the form of the session. Providers shape the delivery details. |

*In all cases agencies had the scope to negotiate session dates, times, duration (between 1-2 hours) and attendance
sites, our fidelity criteria had to cover both standardised and locally adapted intervention components and reconcile potentially disparate adaptations.

3. Responsivity to context. The implementation plan was not fully developed when the trial commenced and was going to incorporate a degree of responsivity to shifting agency priorities, so we needed capacity to adjust our fidelity criteria and data collection methods as the need arose. The complexity of the intervention and of the participating organisations precluded any confident prediction about the essential elements’ validity (would they accurately reflect the intervention theory? would they turn out to be essential?) or even their feasibility (could they be implemented as planned?).

Methods
As a result of these uncertainties, we were unable to pre-determine the content, scope and specificity of the essential elements. Consequently, we judged it necessary to identify provisional essential elements and observe them in the field, using empirical evidence from the process evaluation to revise them as required. Our goal was to critique the construct validity of the essential elements [9] and modify them while simultaneously using them as reliable fidelity indicators.

The mixed-method process evaluation focused on three domains: (a) how the intervention was implemented (fidelity assessment), (b) how people participated in and perceived the intervention, and (c) the contexts that mediated this relationship. As shown in Table 3, qualitative and quantitative data collection methods included purposively sampled semi-structured interviews; direct observation and coding of intervention activities; conversations with the intervention designers, implementers and providers; and participant feedback forms. These are described in detail in the SPIRIT process evaluation protocol [67].

The research group (which comprised the intervention designers, implementation team and process evaluation team working in parallel) used the relatively lengthy intervention period as an opportunity to identify, assess and refine hypothesised essential elements during the trial. This was aided by the multi-agency, stepped wedge design of the

![SPIRIT action framework](https://example.com/spirit-framework.png)
trial which allowed us to monitor the entire intervention in some agencies and still have scope to trial revisions in other agencies. A modified version of this approach could be applied to other trial designs.

The provision of a dedicated process evaluation researcher as part of the wider group enabled the collection of multiple forms of evaluative data from all sites, and iterative conversations with the intervention designers about their conceptualisation of the intervention’s causal pathways. This allowed us to assess the validity of the essential elements using a five-stage process. Stage 1: identify provisional essential elements; stage 2: test provisional essential elements in intervention contexts; stage 3: refine provisional essential elements and develop likely essential elements; stage 4: test likely essential elements in intervention contexts; and stage 5: refine the likely essential elements and develop final essential elements. See Fig. 2 for a visual overview of this process. Each of these stages is now described.

Results
These results overlap with our methods in that we show how process evaluation data collection and analysis was used to critique essential elements. This detail is provided so that the procedure we devised is transparent and replicable.

Stage 1: identifying provisional essential elements
SPIRIT drew on diverse literature and expertise in its design. As shown in Fig. 2, this body of knowledge was distilled by the intervention designers into an action framework (Fig. 1) and a list of change principles (Table 2) [63, 66, 67] which formed the theoretical basis that we attempted to operationalise in response to each intervention session. These sessions were developed by the intervention designers in consultation with agency staff and expert providers.

We could not use SPIRIT’s change principles as our essential elements. Doing so may have been appropriate for a very flexible intervention with minimally specified, non-standardised components [61]. In such a case, fidelity assessment could focus less on specific operationalisations of the change principles and more on if and how the change principles were realised [59]. However, this was not appropriate for SPIRIT which sought a balance of standardisation and flexibility within a menu of predefined components. The process evaluation aimed to report on variation in the delivery and response to each of these components, consequently the change principles were too abstract to be used as indicators for fidelity reporting. Similarly, the action framework, which functioned as our logic model, outlined causal pathways and relationships in relation to individual and organisational capacity building but did not identify techniques. We needed a concrete and observable expression of what was at the heart of these strategies if we were to identify commonalities and differences in implementation that could help interpret the outcomes and inform further interventions.

The approach we devised was to identify potential essential elements inductively. As each session outline became available, the process evaluation team asked three questions. (a) What do the session goals and the planned characteristics of the session tell us about which change principles this session is attempting to utilise? (b) Which of these are likely to be essential to the effectiveness of the session? (c) What would these change principles look like in delivery (how can we operationalise them so that can be measured or fully described)? This produced a list of draft essential elements that we further developed with the SPIRIT designers to accurately describe the elements they believed were essential for that session to be effective. These potential essential elements included session content, key messages, provider characteristics, presentation techniques, activities, and particular attendees and types of participation. At this stage, we consciously trialled many essential elements that we

Table 2 SPIRIT change principles

<p>| Systems framework                                          | • Uses a multi-component approach                                 |
|                                                           | • Maximises interaction between the different components of the intervention |
|                                                           | • Addresses systems, operations, structures and relations          |
|                                                           | • Is flexible in meeting the needs of different agencies           |
| Engagement and ownership                                  | • Engages agencies in owning and driving the program               |
|                                                           | • Is tailored to focus on the agency’s priorities                  |
| Goal setting and feedback                                 | • Provides feedback about current practice                         |
|                                                           | • Provides a clear rationale for change                            |
|                                                           | • Develops agreement about concrete and specific change goals      |
|                                                           | • Monitors and provides feedback about change during the intervention program |
| Interactive skill development                             | • Provides self-education opportunities and access to resources     |
|                                                           | • Recognises the expertise of participants                         |
|                                                           | • Is interactive with a focus on shared reflection and problem solving |
|                                                           | • Provides opportunity for rehearsal and practice                   |
| Leadership, roles and relationships                        | • Uses champions to model and promote the use of evidence from research (including both internal and external champions) |
|                                                           | • Uses credible, dynamic experts as presenters                     |</p>
<table>
<thead>
<tr>
<th>Questions used to critique essential elements</th>
<th>Data sources</th>
<th>Data examples</th>
<th>Data analysis / use</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. When implemented in these contexts, does this provisional / likely essential element realise the change principle(s) that informed its development?</strong></td>
<td>Implementation checklist completed during the delivery of each session</td>
<td>Codes showing whether or not (or to what extent) each essential element was delivered as intended</td>
<td>Collation of codes by session and by agency</td>
</tr>
<tr>
<td></td>
<td>Fieldnotes made during observation of each session</td>
<td>Description of how the essential elements appeared to work or not (e.g. how participants reacted), how they were delivered, any adaptations that took place, any factors that appeared to affect how the intervention was delivered or how people engaged with and responded to it</td>
<td></td>
</tr>
<tr>
<td><strong>2. Is this essential element critical for achieving the session goals? Does anything else appear to be?</strong></td>
<td>Participant feedback forms collected at the end of each session</td>
<td>How participants assessed delivery against quality criteria such as content relevance, provider credibility, and learning outcomes; and their advice for improvements</td>
<td>Descriptive analysis of quantitative data (frequencies, averages and comparisons)</td>
</tr>
<tr>
<td></td>
<td>Transcripts of semi-structured interviews with purposively sampled participants from two phases of interviewing: early in the intervention period and after it</td>
<td>Participant perceptions of the strategies used to effect change: the extent to which they worked and how modifying factors such as work practices, organisational goals, and beliefs about research shaped process effects</td>
<td>Managed using Framework Analysis. Data was synthesised in categories that were identified both inductively from early interviews and a priori based on intervention outcomes and a review of the research utilisation literature</td>
</tr>
<tr>
<td><strong>3. Does this essential element function across all subcomponents and all six trial intervention settings?</strong></td>
<td>Fieldnotes documenting informal conversations with participants following sessions</td>
<td>As above but ad hoc and generally very brief</td>
<td>Data was collated in running memos and, where appropriate, coded thematically using the constant comparative method</td>
</tr>
<tr>
<td></td>
<td>Memos documenting conversations with intervention implementers and providers</td>
<td>Implementers’ views on discrepancies between what was intended and what was delivered. Providers’ accounts of why they ‘went off script’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Memos documenting consultations with the intervention designers</td>
<td>How the designers envisaged the change principles manifesting in intervention sessions</td>
<td></td>
</tr>
</tbody>
</table>
suspected would be collapsed or discarded later. See Additional file 1 for an example.

Devising potential essential elements also required the operationalisation of some relatively abstract overarching concepts. We describe the development of one of these—the concept of quality—in more detail. This is because it is particularly important for ensuring that intervention objectives are achieved [10], yet is neglected in the literature [12, 68].

As per Dusenbury et al.’s definition of quality as ‘the extent to which a provider approaches a theoretical ideal in terms of delivering program content’ ([10]: 244), we conceptualised quality as congruence between (a) the intervention-as-implemented and (b) the intervention theory—in particular, the change principles. The change principles were strongly informed by adult learning theory which provided quality constructs such as: the providers’ content expertise and presentational skills; the extent to which participants found workshops to be interesting, engaging and respectful of their contributions; the relevance and potential usability of the information and ideas provided; and if participants were facilitated to explore how information and ideas might be applied in their work settings [69, 70].

We were able to operationalise some aspects of these quality constructs and so include them as evaluator-coded essential elements (e.g. by devising criteria for ‘content expertise’ and using observations to determine the extent to which information and ideas were discussed in relation to participants’ work). However, because quality is highly situated [12], we considered many aspects would be best assessed by participants themselves. Therefore, items in the participant feedback forms were used to collect information about quality constructs such as content relevance, provider suitability, how engaging the session was and the usefulness of information provided. Quality across the whole program was also considered as part of the semi-structured interviews that were conducted with participants after the intervention.

Interviews focused on capturing the breadth of quality criteria from participants’ perspectives (we were mindful that our notion of quality might not align with theirs) and exploring reasons for their judgement rather than ratings.

Stage 2: testing provisional essential elements in intervention contexts
During the first step of SPIRIT (in which the intervention was fully implemented in two agencies and partially implemented in a further two), the process evaluation team not only monitored adherence to the essential elements but also gathered qualitative and quantitative data that would help us better understand their real world functionality and validity. We conceptualised validity as (1) how well the essential elements embodied and delivered the intervention’s theoretical foundations [6, 9, 71] and (2) the extent to which the essential elements were actually essential in each setting [17] (we were aware that elements which seemed essential in one context might not be so in all contexts and circumstances [13]). Data was collected via observational field notes, checklist coding, post-session memos, participant interviews, participant feedback form ratings and comments, and conversations with providers and implementation team members.

During the concurrent data collection and analysis process, we adopted a stance of ‘naïve curiosity’ in relation to the essential elements, asking ‘What seems to be more or less successful in meeting the goals of each session, and why?’ This enabled us to stay open to potential essential elements that we may have failed to consider prior to the evaluation. For example, we noted early on that participants appeared to engage more with session content and gave more favourable feedback when the provider explicitly recognised the challenges of their work, including having a realistic view of the (limited) role of research within it. When the reverse
was observed (participants disengaging because the provider appeared insensitive to this issue), we concluded this concept was an essential element of the relevant components: ‘Provider demonstrated sensitivity to the ‘real world’ of the agency’s policy/program work.’

To address our concern about validity we also asked ‘How well was the theory underpinning the intervention realised in the delivery of this session?’ and ‘Does each putative essential element appear to be critical for achieving the session goals?’ Data was synthesised in running memos that identified issues to explore in further sessions. Analysis focused on comparing our data with the program logic and, primarily, with the change principles that had been identified as informing each session plan.

Six (often overlapping) challenges to the validity of the essential elements were identified through this inductive process. Essential elements could be (1) redundant—the element was not essential; (2) poorly articulated—unclear, too specific or not specific enough; (3) infeasible—it was not possible to implement the essential element as intended; (4) ineffective—the element did not effectively deliver the change principles; (5) paradoxical—counteracting the goals of the session or the underpinning change principles; or (6) absent or suboptimal—we identified additional or more effective ways of operationalising the change principles. See Table 4 for examples.

Detailed notes were made about the nature of the problem, what interactions affected it (where this was appropriate) and possible solutions that took account of our growing appreciation of contextual constraints and opportunities. Notes included suggestions about where session-specific essential elements could be collapsed and rephrased so that they could be applied across all agencies and intervention components.

Stage 3: refining the provisional essential elements and developing likely essential elements
The process evaluation team used these notes to amend, distil or reject the 50+ provisional essential elements initially used across the intervention into a list of 26 ‘likely’ essential elements. Following consultation with the intervention designers, these were further revised. The list represented a revised way of articulating and evaluating the fidelity of the intervention but did not affect its design or continuing implementation (with the exception of providers who were sent a list of the essential elements and feedback form items prior to their sessions).

In the revision process, we sought to balance the need for more loosely specified essential elements (which the flexible aspects of the intervention design demanded) with the need to clearly describe what the intervention comprised: not only for the purposes of fidelity assessment but also to provide detailed information that would aid transparent reporting of and potential replication of the intervention. We were guided by Century, Rudnick and Freeman’s account of reducing the granularity with which their essential elements were defined and measured [55]. Consequently, essential elements that had been devised for topic specific sessions were articulated at a higher level of abstraction. For example, ‘The provider demonstrated the value of using systematic reviews in policy/program decision-making’ became ‘The value of using research/evaluation in agency work was conveyed’. This was necessary because agencies were able to choose and tailor different sessions from within the same intervention component. So in order to monitor fidelity comparatively across all agencies, the essential elements needed to be applicable to every session. Where agencies were able to choose the topic, content, form and goals of face-to-face sessions, the fidelity assessment no longer specified any of these attributes, only that they must reflect the relevant change principles for that component (e.g. those specifying interactivity, shared problem solving, and recognition of participants’ expertise).

Stage 4: testing ‘likely’ essential elements in intervention contexts
In this stage, we used the likely essential elements in our fidelity assessment data collection and continued using the methods described in stage 2 to collate information about the extent to which they were delivered and to explore their functionality and congruence with the program theory.

Stage 5: developing final essential elements
Several further changes were made in this stage but, with some exceptions, not as a result of additional information gathered in stage 4. Rather the iterative process of refinement allowed us to reflect on details that had been sidelined by more pressing concerns in the previous stages. Having addressed those, we had capacity to focus on less critical amendments and fine tune some essential elements that might otherwise have been considered ‘good enough’. Our final list of essential elements was reduced to 20 items (Table 5). These included several that we considered collapsing but decided to retain separately. For example, is this provider-related element: ‘The provider encouraged participants to contribute to session’ really essential when a participation-related element: ‘Participants contributed to session’ addressed the same concept? Based on empirical evidence from the trial, we concluded it was important to differentiate between (and learn from) what was delivered and how people responded. Our observational data showed that
<table>
<thead>
<tr>
<th>Challenges: the putative essential element was...</th>
<th>Definition</th>
<th>Essential element example</th>
<th>Comments</th>
<th>Suggested response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redundant</td>
<td>The strategy described by the element was not essential</td>
<td>‘The provider encouraged participants to ask questions’</td>
<td>This was unnecessary in discussion-based sessions where participants interacted as co-contributors</td>
<td>Remove this element</td>
</tr>
<tr>
<td>Poorly articulated</td>
<td>The element description was unclear, too specific or not specific enough</td>
<td>‘The session was introduced by a leader (senior person in the agency e.g. CEO, member of executive)’</td>
<td>This failed to capture the many times that less senior staff introduced sessions that were attended by leaders. This essential element was a proxy for visible endorsement/support (modelling) by organisational leaders which we concluded was also achieved when they attended and contributed enthusiastically to the session in other ways</td>
<td>Hone the description so that it accurately captures the essential element</td>
</tr>
<tr>
<td>Infeasible</td>
<td>The essential element described a strategy that was not possible to implement as intended</td>
<td>‘Participants were facilitated to identify one or more change goals’</td>
<td>We found this was achievable only in agencies that had developed a research utilisation reform agenda prior to SPIRIT and felt able to use intervention sessions to discuss their goals openly. Other agencies needed more time and different processes to identify goals</td>
<td>Modify or develop alternative strategies. In some cases, the outcomes themselves may need be modified</td>
</tr>
<tr>
<td>Ineffective</td>
<td>In practice, the strategy described by the essential element did not effectively deliver the change principles</td>
<td>‘The provider had experience presenting to policy/program developers’</td>
<td>This seemed intuitively reasonable as one of several criteria for securing providers with the expertise and credibility stipulated by our change principles, yet there was no correlation between this criterion and our evaluation of session quality or general participant satisfaction feedback</td>
<td>Consider whether this element can simply be removed or if the change principles require further operationalisation to capture an essential aspect of the intervention</td>
</tr>
<tr>
<td>Paradoxical</td>
<td>When implemented, the strategy described by the essential element counteracted the session goals or the change principles</td>
<td>No examples of this were identified</td>
<td>Interventions can have counterintuitive impacts. While the process evaluation identified examples of this in other aspects of the trial, none related specifically to the essential elements</td>
<td>Remove this element and consider possible implications for other parts of the intervention</td>
</tr>
<tr>
<td>Absent or suboptimal</td>
<td>Additional or more effective ways of operationalising the change principles were identified</td>
<td>‘The provider persuasively articulated his/her commitment to using research.’</td>
<td>Despite being briefed to do so, many providers did not articulate their commitment to using research. However, some used case examples that powerfully illustrated the value of research, and facilitated discussion that enabled participants to express it themselves. This strategy was more sophisticated and a better fit with the adult learning orientated change principles that emphasise interactivity, shared reflection and harnessing participant expertise</td>
<td>Introduce absent elements and modify sub-optimally operationalised elements that the essential aspects of the intervention are captured</td>
</tr>
<tr>
<td>Final essential elements</td>
<td>Final scoring of essential element</td>
<td>Activity that provided data for scoring</td>
<td>Intervention components to which essential elements apply</td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------------------</td>
<td>----------------------------------------</td>
<td>----------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>1. Provider had expertise and credentials in the topic/field appropriate to the session</td>
<td>Yes / No</td>
<td>Review of publicly available biographical information and, for no. 1, participant feedback form item</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>2. Provider had experience in presenting to policy/program developers</td>
<td>Yes / No</td>
<td>✓✓ ✓ ✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engagement and facilitation: the methods used to deliver the presentation and encourage participation</td>
<td>Extensive</td>
<td>Direct observation of session delivery</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>3. Non-didactic presentation strategies were used</td>
<td>Extensive</td>
<td>Direct observation of session delivery</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>4. Content was delivered in an engaging manner</td>
<td>Yes / No</td>
<td>✓✓ ✓ ✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. The provider encouraged participants to contribute to session (ask questions, make comments, provide examples, participate in discussion)</td>
<td>Extensive</td>
<td>Direct observation of session delivery</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>6. The provider encouraged participants to discuss how information / learning from the session might be applied in their setting</td>
<td>Extensive</td>
<td>Direct observation of session delivery</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>7. Provider showed respect for participants’ contributions and work</td>
<td>Extensive</td>
<td>Direct observation of session delivery</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>8. Provider demonstrated sensitivity to the ‘real world’ of the agency’s policy/program work</td>
<td>Extensive</td>
<td>Direct observation and participant feedback form item</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>Content: key topics, messages, activities and resources</td>
<td>Aggregated rating across all items specified in session plan: Wholly</td>
<td>Direct observation and multiple participant feedback form items</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>9. Core content outlined in session plan was delivered</td>
<td>Mostly</td>
<td>✓✓ ✓ ✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. The session content was relevant to the agency’s work</td>
<td>Yes</td>
<td>Participant feedback form item</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>11. Where specified in the session plan, provider identified or provided resources that supported or extended learning from the session</td>
<td>Yes / Partially / No / N/A - not specified in plan</td>
<td>Direct observation of session delivery</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>12. The value of using research / evaluation in agency work was conveyed</td>
<td>Yes</td>
<td>Participant feedback form item</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>13. Synthesised data from measures was provided and discussed</td>
<td>Yes</td>
<td>Direct observation of session delivery</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>14. Opportunities to improve use of research were identified</td>
<td>Extensive</td>
<td>Direct observation of session delivery</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>Participation: characteristics of attendees’ interaction and contribution to the session</td>
<td>Numbers and roles of all attendees. Approximate proportion of those targeted</td>
<td>Direct observation and review of data from session ‘sign in sheet’</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>15. Targeted agency staff attended</td>
<td>✓✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5 Overview of SPIRIT's final essential elements: their scoring, how they were monitored and which of the interventions components they applied to (Continued)

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>Direct observation of session delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td>16. A leader (e.g. CEO, member of executive) introduced the session or contributed to it positively in other ways</td>
<td>Yes</td>
<td>No</td>
<td>✓ ✓</td>
</tr>
<tr>
<td>17. Participants contributed to session (asked questions, made comments, participated in discussion)</td>
<td>All</td>
<td>~ ¾</td>
<td>~ ½</td>
</tr>
<tr>
<td>18. Participant contributions included knowledge/examples from their own experience</td>
<td>Extensive</td>
<td>Moderate</td>
<td>Limited</td>
</tr>
<tr>
<td>19. Discussion included how information/learning from the session might be applied in their setting</td>
<td>Extensive</td>
<td>Moderate</td>
<td>Limited</td>
</tr>
<tr>
<td>20. Participants identified one or more agency research-related areas that could benefit from improvement</td>
<td>Yes</td>
<td>No</td>
<td>Direct observation of session delivery</td>
</tr>
</tbody>
</table>

*Essential elements are one type of fidelity criteria. Other fidelity measures concerning frequency, duration, coverage, etc., plus participants’ perspectives, were collected for each session but are not shown on this table.*
in most sessions the providers’ actions appeared to shape the levels and types of participation, but this was not always the case. Also, because providers were given a loosely specified briefing regarding delivery techniques, as befitted the senior experts who were recruited, we felt it helpful to retain the item for instructional purposes.

**Scoring the essential elements**

Not all fidelity criteria can be assessed in the same manner [9]. Structural items such as participant attendance and the number, type and duration of sessions are easily observed and can usually be captured numerically. However, process items (which may be more significant in terms of intervention effects [9]) such as presentation styles, types of participation and overall quality tend to be more descriptive and usually require context-sensitive qualitative assessment, especially direct observation [9, 19, 62]. Most of our essential elements were processual so we found that their inclusion in the fidelity assessment required that they be monitored not only in terms of whether they were delivered, but the extent to which they were delivered and how this was done. Our aim was to devise a pragmatic method of standardising observations across sites that could accommodate local adaptation and extensive data collection.

We made three primary adjustments to the scoring as a result of the testing. First, we rejected dichotomised scoring on many items in favour of an ordinal scale. Not surprisingly, we found the yes/no format we trialled too reductive for the complex processes we were observing. We also trialled several five-point scales (as recommended by Bond et al. [21]) but settled on a four-point descriptive scale of extensive/moderate/limited/not at all as providing the necessary breadth and precision for our purposes. The definitions that specified the conditions under which each score was applicable were refined in consultation with the intervention designers and the scale was tested in each agency by two members of the team. All coding was supplemented with descriptive notes.

Second, we developed a scale that could be applied to each customised session (workshop, symposium, etc.) and would thereby enable us to compare session content scores across the whole trial. Content was considered to be the aspects of the session that the participating agency had specifically requested. Depending on the nature of the session and the level of detail each agency chose to specify, this content varied tremendously from concrete deliverables (e.g. an example of a systematic review was provided) to relatively abstract processes and concepts (e.g. ethical challenges were explored interactively). The number of content items also varied from between three to eight. We kept the yes/no score for each individual item and simply aggregated these using a scale of wholly/modestly/about half/limited/not at all for each session. This allowed us to compare the delivery of varied content across all sessions and sites without the requirement for a consistent number of items.

Third, we concluded that we had been unsuccessful in finding semi-objective generalisable ways of scoring certain quality concepts (e.g. Was the presentation engaging? Was the content relevant?). We decided to rely entirely on participant feedback to score these essential elements. See Table 5 for an overview of the final scoring.

We had sufficient data (checklists, descriptive notes, memos and audio recordings) from the intervention implementation in stage 1 to apply these new codes retrospectively to the sessions that informed them.

**‘Prohibited’ elements**

During the trial, we eschewed the concept of ‘prohibited’ [9] or ‘forbidden’ elements [72], but when reviewing the data for stage 5 revisions, we concluded that they could have provided clearer guidance for our providers about the intervention’s underpinning principles. These providers were experts in their field but newcomers to SPIRIT. Despite receiving the essential elements for their sessions in advance, many appeared to apply them selectively. Based on participant feedback and our observations, the following guidance may have helped providers avoid the most common pitfalls:

*To be avoided:*

- **Talking down to participants.** In particular, failure to recognise their expertise and the complexity of their work.
- **Talking at participants.** Didactic presentations should be interspersed with case examples, activities, discussion, etc. Invite questions, ask participants about their views and experiences, and encourage debate.
- **Reliance on data/cases from other fields.** When information is highly relevant it is more applicable. Where possible, use case examples from the agency’s own work. We can provide assistance with this.
- **Squeezing out time for discussion.** We conceptualise discussion as a primary mechanism for helping participants integrate new knowledge and think about how it might be applied in their contexts.

We did not trial this guidance partly because it would have radically changed the provider briefing protocol and partly because of the potential to alienate eminent highly skilled professionals with such censorious (and potentially patronising) guidance. However, we believe that our methods for assessing essential elements, combined with sensitive consultation with the providers, would
glean valuable information about the appropriateness and utility of such an approach. Although this paper concentrates on critiquing and revising essential elements in situ as a means of improving validity in novel contextualised trials, where threats to validity can be identified in advance they should be addressed before the intervention is underway.

Discussion

Identifying an intervention’s essential elements and monitoring them via fidelity assessment is critical for understanding how the intervention worked or why it did not work. Yet, there is uncertainty about how to do this, particularly for novel contextualised interventions (i.e. interventions that blend theories pragmatically and which are designed to be flexible and at least partially responsive to local conditions) [8–10, 20, 55]. How do we determine which elements of such interventions are genuinely essential to their effectiveness? And how do we ensure they are valid indicators of the intervention theory [6, 12, 14]? When attempting to answer these questions we found little practical guidance in the literature and encountered paradigmatic differences and ambiguous terminology. For example, what we call essential elements [10, 56] are also known as essential functions [59], essential components [12], essential ingredients [62], active ingredients [6, 7, 11], critical ingredients [21], critical components [55] and core components [23, 36]. More importantly, they are not always referring to the same phenomenon and they differ greatly in terms of their relationship to the intervention’s theoretical underpinnings. Some refer theoretically to intervention activities [12], others to theoretical functions [59]; some use the term to include the breadth of fidelity criteria (e.g. intensity and reach) [20], while others limit it to carefully mapped and validated indicators of theory-based models [73] or recommendations [17].

Meanwhile, the perceived value of assessing standardised interventions using universal fidelity criteria is declining. The growth of contextualised interventions mirrors increasing recognition of the complexity of the dynamic real world systems in which they are implemented, and the idiosyncratic and unintended ways that interventions and their context can change one another [41, 49, 59, 74]. The need to figure out what fidelity means in such interventions, and to devise methods for identifying and monitoring elements that are genuinely essential, is more pressing than ever.

In this paper, we describe a novel exploratory incremental test-and-refine process devised to strengthen the validity of a contextualised intervention’s essential elements. This pragmatic approach enabled us to collect fidelity data throughout the trial (despite uncertainty about what the intervention would look like when implemented in each setting), while also assessing how well the intervention’s real world delivery aligned with the theoretical principles that underpinned its design. The literature provides advice for articulating factual, precise and targeted fidelity criteria prior to the intervention e.g. [21] but to ensure our essential elements were valid we needed to attend to the interplay of the intervention theory and design with the intervention settings, providers and participants. This was best done empirically in the context of the trial.

Although we monitored implementation fidelity, our methods focused on understanding the intervention’s theoretical fidelity because, as Hawe argues, ‘Fidelity resides in the theory of the change process, rather than in any particular technology, component, or delivery channel per se. Thus, the role and meaning behind a particular component, rather than its face value, are what matter’ ([75]: 313).

Identifying the appropriate level of specificity was a critical aspect of determining the essential elements’ validity. Overarchingly, we moved from a tightly specified approach to one that was more loosely defined, better reflecting the intervention’s scope for expert providers to shape activities, and for tailoring to individual sites. We knew that session-specific essential elements would need to be distilled into higher order items that covered whole components of the intervention, but testing the functionality and theoretical congruence of a wide variety of provisional essential elements in multiple sessions and sites enabled us to explore a breadth of possibilities about what mattered and why, increase our understanding of which intervention elements genuinely appeared to be essential, and experiment with how best to articulate and score them. One outcome of this was to increase the extent to which participant feedback was used to measure quality indicators. This approach accords with calls in fidelity assessment, and in research and evaluation more broadly, to use loosely specified evaluation methods that support local adaptation and which recognise that change processes in complex systems are unpredictable and are often best assessed by those receiving the intervention [7, 38, 58, 59]. While none of the process evaluation data, including the evolving fidelity assessment described in this paper, was fed back into the design or implementation of the intervention during this trial, our approach has potential to contribute formatively to developmental evaluations that shape the intervention during its delivery [52].

Our fidelity data will be analysed in relation to participants’ feedback form ratings for each intervention session. We anticipate that sessions with higher implementation fidelity will receive a higher overall score and more favourable free text responses. It will not be possible to disentangle the implications of fidelity
results for individual sessions or components when analysing intervention outcomes as they are thought to function interdependently, but our data will tell us the extent to which the operational and theoretical aspects of the SPIRIT intervention were delivered in each agency. This, in turn, will help us interpret the observed effects of the overall intervention-as-delivered on outcomes.

The use of mixed data collection methods and sources (triangulation), including direct observation and participant feedback, strengthened the rigour of this work [9, 19, 21, 62]. However, the final recursive loop (stages 4 and 5 as described in the ‘Methods’ section) could have been avoided if we had scrutinised all the essential elements with equal emphasis in earlier steps rather than focusing on those with evident problems.

We note that this approach would not be appropriate for all interventions. Given that the modifications mostly either collapsed essential elements or articulated them at a less granular level, we were able to use the data gathered during earlier phases of implementation to apply the modified elements and codes to the sessions that informed them. However, where essential elements are revised to become more granular (as might be the case in standardised programs where highly specified techniques are being honed), our records would not have contained sufficient detail with which to apply codes retrospectively.

There are other limitations. Our lack of independence as members of the wider study team may have affected our ability to observe the intervention implementation dispassionately and, as is always the case, our theoretical and disciplinary allegiances may have skewed what we noticed and how we assessed it. Lastly, what we observed was situational: shaped by the complex interaction between the intervention theory and structure, delivery by multiple providers, diverse participants and distinct organisational contexts, all at particular time points. So, while we believe we have identified elements that are at the heart of the intervention theory, we cannot claim that they will necessarily have equal functionality and validity in all settings and circumstances, particularly where they are expressed with greater specificity [65, 68]. We have, however, honed a list of essential elements that appear to be valid in the context of this trial, and which may provide a starting point for others for interventions similar to SPIRIT.

Conclusion
This paper describes the difficulties in identifying the essential elements of a contextualised intervention (i.e. an intervention that is informed by composite social and psychological theories and which incorporates standardised and flexible components in order to maximise effectiveness in complex settings). A worked example of an approach for critiquing the validity of essential elements is provided, including a demonstration of how they can be refined during a trial without compromising the fidelity assessment. This process takes intervention evaluators closer to making theoretically and contextually sensitive decisions upon which to base fidelity assessments in trials of contextualised interventions.

Additional file

Additional file 1: Example of how essential elements changed during SPIRIT. (DOCX 30 kb)

Competing interests
SPIRIT is funded as part of the Centre for Informing Policy in Health with Evidence from Research (CIPHER), an Australian National Health and Medical Research Council Centre for Research Excellence (APP1001435) and administered by the Sax Institute. The Sax Institute receives a grant from the NSW Ministry of Health. The Australasian Cochrane Centre is funded by the Australian Government through the National Health and Medical Research Council (NHMRC). AH is supported by an NHMRC Public Health and Health Services Postgraduate Research Scholarship (1093096).

Authors’ contributions
AH led the design and conduct of the fidelity assessment and drafted the manuscript. SB and GG contributed to the design and ongoing oversight of this work as part of the SPIRIT process evaluation. AW supervised this work. SR and the CIPHER team investigators conceived of the SPIRIT study. PB helped draft the manuscript. All named authors contributed substantially to and approved the final manuscript.

Acknowledgements
We wish to thank the people and organisations participating in SPIRIT. SPIRIT is being conducted by the CIPHER Centre for Research Excellence. CIPHER is a joint project of the Sax Institute; Australasian Cochrane Centre, Monash University; University of Newcastle; University of New South Wales; Research Unit for Research Utilisation, University of St Andrews and University of Edinburgh; Australian National University; and University of South Australia. Thanks also to the members of the CIPHER team who provided valued support and feedback during the SPIRIT process evaluation, particularly: Stacy Carter, Denise O’Connor, Carmen Huckel Schneider and Tari Turner. Lastly, thanks to the reviewers whose comments improved our manuscript.

Author details
1Sax Institute, Level 13 Building 10, 235 Jones Street, Ultimo, NSW 2007, Australia. 2School of Public Health, University of Sydney, Edward Ford Building (A27), Fisher Road, Sydney, NSW 2006, Australia. 3Australasian Cochrane Centre, School of Public Health and Preventive Medicine, Monash University, The Alfred Centre, 99 Commercial Road, Melbourne, VIC 3004, Australia. 4School of Medicine, University of Notre Dame, 160 Oxford Street, Darlinghurst, NSW 2010, Australia. 5Faculty of Health Sciences, University of Sydney, 75 East Street, Lidcombe, NSW 2141, Australia. 6Centre for Medical Psychology and Evidence-based Decision-making, Level 6, Chris O’Brien Lifehouse (C392), University of Sydney, Sydney, NSW 2006, Australia.

Received: 22 July 2015 Accepted: 29 January 2016
Published online: 24 February 2016

References
1. Grimshaw JM, Zwarenstein M, Tetroe JM, Godin G, Graham ID, Lemyre L, et al. Looking inside the black box: a theory-based process evaluation alongside a randomised controlled trial of printed educational materials (the Ontario printed educational message, OPEM) to improve referral and


Submit your next manuscript to BioMed Central and we will help you at every step:

• We accept pre-submission inquiries
• Our selector tool helps you to find the most relevant journal
• We provide round the clock customer support
• Convenient online submission
• Thorough peer review
• Inclusion in PubMed and all major indexing services
• Maximum visibility for your research

Submit your manuscript at www.biomedcentral.com/submit
Appendix 11. An example of how selected essential elements changed during SPIRIT

This table shows the evolution of essential elements for the ‘appraising research’ symposium

<table>
<thead>
<tr>
<th>Providers</th>
<th>‘Provisional’ essential elements (session specific)</th>
<th>‘Likely’ essential elements (applicable to all sessions within this component)</th>
<th>Final essential elements (applicable to all sessions within this component)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The provider had:</td>
<td>• Research credentials at an equivalent level or higher than the attendees • Experience in presenting to policy/program developers</td>
<td>The provider had:</td>
<td>• Expertise and credentials in the topic/field appropriate to the session • Experience in presenting to policy/program developers</td>
</tr>
<tr>
<td></td>
<td>A senior manager in the agency:</td>
<td>The presenter:</td>
<td>A leader (e.g. CEO, member of executive) introduced the session</td>
</tr>
<tr>
<td></td>
<td>• Introduced the session • Explained that audit had identified a need/desire for this symposium • Participated in the session</td>
<td>• Acknowledged the sophistication and complexity of policy/program work • Acknowledged the challenges and constraints of using research in policy/program work • Showed respect for participants’ contributions • Encouraged and facilitated: - group discussion - questions from participants - contributions from participants that addressed challenges and/or questions - examples from participants’ practice experience - comments from participants that demonstrated their skills and knowledge • 1 &gt; appraisal examples were worked through in small groups and fed back, OR in larger group with participants providing feedback</td>
<td>The provider: • encouraged participants to ask questions • encouraged participants to discuss one or more aspects of the topic • encouraged participants to discuss how the information/learning might be applied in their setting • showed respect for participants’ contributions • demonstrated sensitivity to the ‘real world’ of policy/program work</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Non-didactic teaching strategies were used • Content was delivered in an engaging manner</td>
<td>• Non-didactic teaching strategies were used • Content was delivered in an engaging manner</td>
</tr>
</tbody>
</table>

Content was delivered in an engaging manner
<table>
<thead>
<tr>
<th>Session content</th>
<th>Participation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The presenter:</strong></td>
<td><strong>Targeted agency staff attended</strong></td>
</tr>
<tr>
<td>- described evidence appraisal</td>
<td>- Targeted agency staff attended</td>
</tr>
<tr>
<td>- persuasively articulated the value of appraising evidence</td>
<td>- Participants asked questions</td>
</tr>
<tr>
<td>- facilitated discussion about the role that appraisal can &amp; should play in policy/program development, including acknowledgement of:</td>
<td>- Participants contributed to discussion</td>
</tr>
<tr>
<td>- time pressures</td>
<td>- Participant contributions included knowledge/examples from their own experience</td>
</tr>
<tr>
<td>- pragmatism (policy goals may be different to academic goals)</td>
<td>- Discussion included how information/learning from the session might be applied in their setting</td>
</tr>
<tr>
<td>- described different types of evidence (e.g. single studies, reviews, meta-reviews, evidence based recommendations, summaries for policy and practice)</td>
<td></td>
</tr>
<tr>
<td>- described the value and uses of evidence types, including types of Qs they can answer</td>
<td></td>
</tr>
<tr>
<td>- used examples to illustrate evidence types</td>
<td></td>
</tr>
<tr>
<td>- described how to appraise evidence, including:</td>
<td></td>
</tr>
<tr>
<td>- what makes a review reliable</td>
<td></td>
</tr>
<tr>
<td>- assessing applicability: generalisability, relevance, transfer to local settings</td>
<td></td>
</tr>
<tr>
<td>- described useful evidence sources, including:</td>
<td></td>
</tr>
<tr>
<td>- where to find reviews and summaries</td>
<td></td>
</tr>
<tr>
<td>- what different repositories cover</td>
<td></td>
</tr>
<tr>
<td>- described useful tools/resources, including:</td>
<td></td>
</tr>
<tr>
<td>- where to find info to guide appraisal of individual studies</td>
<td></td>
</tr>
<tr>
<td>- where to find info to guide own evidence reviews</td>
<td></td>
</tr>
<tr>
<td>Content was relevant to the agency’s work</td>
<td></td>
</tr>
</tbody>
</table>

* The core content for this session comprised four items: 1. What is a systematic review? 2. Where can I find systematic reviews? 3. Methods for appraising systematic reviews, and 4. Practical exercise in which participants work through the appraisal of a systematic review
Appendix 12. Chapter Eight Manuscript

The pivotal position of ‘liaison people’: facilitating a research utilisation intervention in policy agencies
The pivotal position of ‘liaison people’: facilitating a research utilisation intervention in policy agencies

Abby Haynes, Sax Institute and University of Sydney, abby.haynes@saxinstitute.org.au;
Phyllis Butow, University of Sydney, phyllis.butow@sydney.edu.au;
Sue Brennan, Monash University, sue.brennan@monash.edu;
Anna Williamson, Sax Institute, anna.williamson@saxinstitute.org.au;
Sally Redman, Sax Institute, sally.redman@saxinstitute.org.au;
Stacy Carter, University of Sydney, stacy.carter@sydney.edu.au;
Gisselle Gallego, The University of Notre Dame, Sydney and Umeå University, gisselle.gallego@nd.edu.au;
Sian Rudge, Sax Institute, Sian.Rudge@saxinstitute.org.au

This paper explores the enormous variation in views, championing behaviours and impacts of liaison people: staff nominated to facilitate, tailor and promote SPIRIT (a research utilisation intervention trial in six Australian health policy agencies). Liaison people made cost/benefit analyses: they weighed the value of participation against its risks and demands in the context of organisational goals, knowledge utilisation norms, epistemology and leadership support. There was a degree of self-fulfilment (organisations got what they put in), but SPIRIT could not always be tailored to address local knowledge needs. We present nine propositions for identifying and supporting liaison people in similar interventions.

Keywords: research utilisation • faciliation • championing • process evaluation

Introduction

Externally designed and implemented organisational change interventions are thought to have a greater chance of success when they are supported by one or more internal staff members acting as facilitators (Greenhalgh et al, 2004). Such facilitators often manage the administrative tasks associated with an intervention and may be involved in recruitment, consent processes and/or data collection. More importantly, they are social mediators of the ideas and processes central to the intervention. This may involve formal activities such as presentations at staff meetings, but is likely to include ad hoc negotiation and interpretive communication with diverse colleagues and with those implementing the intervention. Thus facilitators are expected to function as persuasive advocates and mediators, using their interpersonal skills and institutional knowledge to deliver and, where necessary, reframe interventions to maximise their success.
In this paper we build on existing knowledge by describing the attributes, perceptions, contexts and associated behaviours of the facilitators – known as liaison people – of a novel complex trial that was designed to increase the use of research in health policy agencies (CIPHER Investigators, 2014). We demonstrate that the liaison people (LPs) functioned as critical mediators with profound impacts on how the intervention was shaped and received in each site. We develop propositions from our analysis that provide guidance about how to identify and support LPs (or related functions) in similar interventions. But first, we present an overview of the key roles and characteristics of intervention facilitators in general, and then describe the intervention trial that our LPs were facilitating.

**Characteristics of internal intervention facilitators**

Intervention facilitators are conceptualised in many ways, but the literature draws attention to three predominant types: champions, opinion leaders and boundary spanners. The terms are not mutually exclusive and are often used interchangeably and/or ambiguously (Greenhalgh et al, 2004; Williams, 2011), but they denote specific attributes and functions with implications for how change agents are identified, supported and utilised. (Greenhalgh et al, 2004; Thompson et al, 2006)

**Champions** are internal employees who advocate for organisational change initiatives. Their function is to capture attention and counter indifference by connecting the intervention with organisational goals and values. Champions articulate their vision of the intervention and demonstrate personal commitment to it (Hendy and Barlow, 2012; Howell and Boies, 2004). This involves risk as the characteristics of the intervention, including its failure or success, will be associated with their judgement and prestige (Thompson et al, 2006). The literature describes champions variously as people who emerge spontaneously during a new initiative (Hendy and Barlow, 2012; Howell and Boies, 2004; Markham, 1998), or respond to a ‘champion call’, or are purposefully recruited (Hammond et al, 2011; Ploeg et al, 2010). Given their need to be genuinely enthusiastic and to be perceived by colleagues as authentic, some argue that champions should not be formally appointed (Howell and Boies, 2004). Championing tactics vary (Greenhalgh et al, 2004) and are powerfully mediated by interpersonal and contextual factors (Locock et al, 2001). This makes it hard to build champions into standardised implementation planning.

**Opinion leaders** are ‘able to influence informally other individuals’ attitudes or overt behaviour in a desired way with relative frequency’ (Rogers, 2003). Although opinion leaders may mobilise members of an organisation through their expert authority or status (Damschroder et al, 2009), they can also be ‘near-peers’: competent and knowledgeable colleagues who have influence partly because they are seen to share the same frames of reference (Locock et al, 2001; Rogers, 2003). Opinion leadership is targeted and topic-specific, so different opinion leaders may be required for different types and stages of a change process. Thus someone who is strongly influential in one setting under particular circumstances may find their views dismissed in other settings, or under different conditions (Grimshaw et al, 2006).

**Boundary spanners** link people, sectors, interests and perceptions (Williams, 2002). Their strong external relationships expose them to ideas in the broader environment so they may be more open than other staff to new ways of doing things (Greenhalgh et al, 2004). They can support interventions by building coalitions and bridging gaps
in understanding between the organisation and those implementing the programme (Williams, 2002). Such gaps are often exacerbated by lack of disciplinary or industry knowledge: a common concern in researcher/policymaker relationships (Caplan, 1979). Unlike opinion leaders and champions, the role of the boundary spanner is often formalised.

Effective knowledge brokers (those who facilitate exchange between producers and users of knowledge) (Lomas, 2007) possess attributes of champions, opinion leaders and boundary spanners. Knowledge brokers support research-informed policy and practice through knowledge management, linkage and exchange, and/or capacity development, requiring credibility, influence, and the technical and communicative expertise necessary to advance knowledge initiatives within and across complex organisational systems. (Conklin et al, 2013; Traynor et al, 2014; Ward et al, 2009)

Many of these characteristics also resonate with Kingdon’s (2003) concept of policy entrepreneurs: well-connected advocates who drive change at a macro level (rather than at the organisational level). They leverage policy opportunities by linking different facets of the political system (aspects of boundary spanning); and combine technical expertise, influential rhetorical skills and political savvy with tenacity and a willingness to devote substantial time and energy to the enterprise (aspects of opinion leadership and championing) (Kingdon, 2003).

Common to all these functions is the centrality of complex social processes (Conklin et al, 2013; Oborn et al, 2011; Thompson et al, 2006). Key individuals can influence organisational change, but ultimately it is negotiated through consultation and comparison with peers (Brown and Duguid, 2001; Weick, 1995). Thus the attributes of successful facilitators can only be understood in relation to their context. Greenhalgh and colleagues found that champions were a key determinant of organisational innovation, but that ‘no amount of empirical research will provide a simple recipe for how champions should behave that is independent of the nature of the innovation, the organizational setting, the sociopolitical context, and so on’ (2004, 615).

Attempts to establish criteria for opinion leaders are similarly confounded: ‘What makes someone a credible and influential authority is derived not just from their own personality and skills and the dynamic of their relationship with other individuals, but also from other context-specific factors’ (Locock et al, 2001, 745). Those developing the concept of facilitation concur, arguing that facilitators require a toolkit of skills and attributes that can be wielded for different purposes and contexts, but that their most critical expertise may be in fully grasping the requirements of specific circumstances and responding flexibly. Thus high quality facilitation is that which is most appropriate to the needs of a particular change situation (Harvey et al, 2002; Wilkinson and Frost, 2015).

**The SPIRIT study and the LPs who supported it**

As mentioned, this paper focuses on ‘liaison people’, internal organisational staff who were nominated to assist with the implementation of Supporting Policy In health with Research: an Intervention Trial (SPIRIT). Six health policy agencies in Sydney, Australia participated in SPIRIT over a 30-month duration. Further details are provided elsewhere (CIPHER Investigators, 2014; Haynes et al, 2014; Haynes et al, 2016; Makkar et al, 2016; Redman et al, 2015).
SPIRIT’s year-long intervention was designed to increase the use of research by staff in health policy agencies. Its components included locally-tailored educational workshops; structured dialogues with experts in research, policy and knowledge brokering; leadership forums focusing on organisational change; the provision of targeted research products and resources; and access to an online information portal. SPIRIT used a stepped wedge cluster randomised trial design in which the intervention was implemented sequentially, with agencies randomly allocated to the time period in which they received the intervention. Outcomes were measured at baseline (prior to any of the six sites receiving the intervention), then at six-monthly intervals, using structured interviews and a self-reported online survey. An in-depth mixed method process evaluation monitored fidelity and explored the interaction between the intervention, participants and contexts.

The SPIRIT investigators initially considered employing a member of staff within each agency part-time to act as an LP. This would have recompensed the LP for their contribution to the study and potentially increased accountability and effort. However, policy colleagues advised that it would be hard to identify staff who would be suitable for (and willing to take on) this dual role, and that shared management would be problematic. Consequently, a more agency-driven approach was used to maximise local ownership of the intervention: the CEO or equivalent in each agency was asked to appoint a suitable member of staff who would act as the LP. This appointment was a requirement of participation in SPIRIT but, due to the diversity of these agencies, there was no stipulation about what attributes the LP should have other than the ability to assist with a range of administrative, decision-making and promotional activities related to the trial.

LPs were provided with a ‘Liaison Person Manual’ that detailed their responsibilities and timeframes (see Table 1), and attended a briefing teleconference with the lead investigator prior to the trial. It was hoped that LPs would assist in maximising awareness and enthusiasm about SPIRIT, as well as ensure it ran smoothly.

Research questions

While we agree that there is no simple recipe for successfully championing an intervention, we argue that understanding critical aspects of the interplay between personal attributes, views, behaviours, context and the nature of the intervention is possible and can help in the selection and support of effective facilitators in similar interventions. Hence, we attempt to answer four inter-related questions:

1. What were the professional characteristics of the people who acted as SPIRIT’s LP and how did these affect engagement with and perceptions of the study (process effects)?
2. How did LPs perceive and promote SPIRIT and with what process effects?
3. To what extent did the LPs operate as champions, opinion leaders and/or boundary spanners?
4. How can we explain the variation in how LPs perceived and promoted SPIRIT? Including (a) What was the role of organisational leaders? and (b) What was the role of organisational context?
<table>
<thead>
<tr>
<th>Phase of study</th>
<th>Task focus</th>
<th>Details</th>
<th>Timing / frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-intervention phase</strong></td>
<td>Liaison person briefing</td>
<td>Attend a teleconference with other LPs in which SPIRIT and the LP role are detailed</td>
<td>A one-off teleconference prior to the intervention for LPs already in place</td>
</tr>
<tr>
<td></td>
<td>SPIRIT introductory session</td>
<td>Attend and participate in the introductory session in which the study (and LP role) is explained to staff</td>
<td>A one-off hour-long session preceding the intervention</td>
</tr>
<tr>
<td><strong>Intervention phase (over 12 months starting at 0, 6 or 12 months from the commencement of SPIRIT)</strong></td>
<td>Selection of intervention components</td>
<td>Consult with colleague / leaders as required to identify optimal components for agency needs</td>
<td>On-going: starting after the agency receives its audit feedback and finishing when all components are selected</td>
</tr>
<tr>
<td></td>
<td>Identification of agency interests and priorities</td>
<td>Consult with colleague / leaders as required to identify topics, content and providers that will best address agency needs</td>
<td>On-going: starting after the agency receives its audit feedback and finishing when all options are agreed</td>
</tr>
<tr>
<td></td>
<td>Intervention activities</td>
<td>Schedule and book resources for intervention activities</td>
<td>Periodically as required over the 12-month intervention period</td>
</tr>
<tr>
<td></td>
<td>Invitations and reminders</td>
<td>Invite colleagues to participate in intervention activities</td>
<td>Periodically as required over the 12-month intervention period</td>
</tr>
<tr>
<td><strong>Data collection (over 36 months starting at commencement of SPIRIT)</strong></td>
<td>Identification of documents and participants for outcome measures (measures are collected every 6 months over 30 months)</td>
<td>Develop an initial list of invitees and contact details for the online survey based on eligibility criteria</td>
<td>The list is developed before measurement point 1, then updated before each of measurement points 2-6</td>
</tr>
<tr>
<td></td>
<td>Invitations and reminders</td>
<td>Nominate four 'best practice' documents and provide details of the people who developed them</td>
<td>Every six months for six measurement periods</td>
</tr>
<tr>
<td></td>
<td>Process evaluation interviews</td>
<td>Nominate a senior member of staff to be interviewed</td>
<td>Every six months for six measurement periods</td>
</tr>
<tr>
<td></td>
<td>Invitations and reminders</td>
<td>Send emails to all eligible staff inviting them to participate in the online survey, followed by two reminder emails</td>
<td>Every six months for six measurement periods</td>
</tr>
<tr>
<td></td>
<td>Process evaluation interviews</td>
<td>Participate in one interview early in the intervention and one following it</td>
<td>Interviews last up to one hour and take place approximately 11 months apart</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>Other liaison tasks as required (not specified in manual)</td>
<td>Advocate for SPIRIT, be a resource for colleagues, provide advice to SPIRIT team, act as a communication link between the agency staff and the SPIRIT team</td>
<td>Ongoing over the duration of the study (36 months), but likely to be more intense during the 12-month intervention period</td>
</tr>
</tbody>
</table>
Methods

In this paper we report data from the SPIRIT process evaluation (Haynes et al, 2014). Primary data collection methods were: semi-structured interviews with the LPs and purposively sampled staff in each of the six agencies; observations of intervention activities (most of which were attended by LPs); and conversations with study staff who were interacting with LPs during the trial (Table 2). Analytic memos written after each data collection event were an additional data source.

LPs were interviewed twice: early in the intervention and post-intervention. Early interview questions addressed: the LP’s work role and tenure, their views about agency research use, how they came to be the LP, initial impressions of SPIRIT, and predictions for how the intervention would be received in their agency. Post-intervention interviews focused on: their experience of acting as the LP, challenges and benefits, how they tackled the LP tasks, factors that affected engagement, any non-SPIRIT activities affecting organisational research use, support internally and by the SPIRIT team, their views of SPIRIT and any feedback about colleagues’ views, any impacts, and improvement advice. In the post-intervention interviews other staff were asked:

The people who took on the role of facilitating SPIRIT in each organisation were quite diverse. In your organisation X acted in that role. How do you think her/his position here or the way she/he approached the tasks involved in facilitating SPIRIT might have affected how people engaged with it?’

Prompts were used to explore participants’ views about the attributes, behaviours and impacts of their LP in more depth.

Interview data were managed in NVivo® (QSR International Pty Ltd, 2012) using Framework Analysis (Ritchie and Lewis, 2003). This allowed us to summarise and categorise the critical dimensions of the data while maintaining links to the verbatim transcripts. Categories were derived from (1) a priori considerations such as the role of organisational leadership and LPs’ characteristics, and (2) constructs developed inductively from the data such as LPs’ perceptions of intervention flexibility and how they integrated LP tasks into their daily work. A later round of analysis was guided by further concepts from the literature, coding for instances of LPs acting as champions, opinion leaders and/or boundary spanners. Observational and memo data was synthesised into schematic case studies which were structured to allow cross-organisational comparison of key dimensions. During analysis, the LP-related interview data was reviewed iteratively against the case studies to contextualise perceptions, relationships and experiences. Data collection and analysis was concurrent, founded on the method of constant comparison where data is iteratively sought and scrutinised in order to develop, nuance and counter emerging hypotheses and explanations (Boeije, 2002). Synthesised LP data and emerging interpretations were reviewed by a small team of multidisciplinary investigators who contributed regularly to the process evaluation work. Later analysis was reviewed by members of the SPIRIT implementation team in order to identify any inaccuracies, and so we could consider their views.

Draft findings were sent to the six primary LPs, that is, the people who acted as LP for the majority of the intervention in their agency. They were asked to comment on the reasonableness of the findings and to inform us of any other views they wished
Table 2: SPIRIT process evaluation data collection details

<table>
<thead>
<tr>
<th>Data source</th>
<th>Data collection method</th>
<th>Data</th>
<th>Timing</th>
<th>Focus of data collection</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Liaison people</strong></td>
<td>Semi-structured interviews</td>
<td>Digital recordings, transcripts and memos</td>
<td>During the early phase of the intervention</td>
<td>How they became the LP? Actions and experience of the role to date. Initial views of SPIRIT and predictions re staff engagement. Soon after the intervention ended</td>
</tr>
<tr>
<td></td>
<td>Informal telephone calls and emails</td>
<td>Notes and email data (summarised in memos)</td>
<td>Throughout the trial</td>
<td>Views / concerns / further reflections about SPIRIT or contextual factors.</td>
</tr>
<tr>
<td><strong>General participants</strong></td>
<td>Semi-structured interviews (purposively sampled)</td>
<td>Digital recordings, transcripts and memos</td>
<td>During the early phase of the intervention</td>
<td>Organisational culture and context, initial views of SPIRIT, LP role in early implementation. Soon after the intervention</td>
</tr>
<tr>
<td><strong>Intervention sessions</strong></td>
<td>Observations and checklist completion</td>
<td>Digital recordings, fieldnotes, checklist codes and memos</td>
<td>Throughout the intervention phase of the trial</td>
<td>Documenting intervention delivery, describing participation and interactions in each agency, including the role of LPs.</td>
</tr>
<tr>
<td><strong>SPIRIT staff</strong></td>
<td>Interviews / structured conversations</td>
<td>Fieldnotes and memos</td>
<td>During engagement phase and after mid-intervention feedback</td>
<td>Any information about views and activities of executive staff and LPs that would help to explain interaction with and impacts of SPIRIT.</td>
</tr>
<tr>
<td></td>
<td>SPIRIT staff meetings and ad hoc conversations</td>
<td>Memos</td>
<td>Throughout trial</td>
<td>Any information about views and activities of LPs and other agency staff that would help to explain interaction with and impacts of SPIRIT.</td>
</tr>
<tr>
<td></td>
<td>Collation of emails from LPs copied to SPIRIT staff</td>
<td>Emails (summarised in memos)</td>
<td>Throughout the trial</td>
<td>Verification of LPs’ email communications to staff in their agency about different aspects of the trial (this information was not received consistently so it is not strictly comparable).</td>
</tr>
</tbody>
</table>
us to consider. We explained that their opinions would be considered and included in the resulting paper, but would not necessarily alter our interpretations. Our purpose was to: (a) provide the primary LPs with an opportunity to contribute to the depiction of LPs in their agency; (b) re-examine our interpretations in the light of potentially challenging insider perspectives; and (c) provide additional data with which readers could critically assess our findings (Locke and Ramakrishna Velamuri, 2009). Unlike conventional member checking this was not an attempt to validate our findings – people may have quite different though equally valid views of the same issues (Sandelowski, 1993). All six LPs responded. Their views, including an overview of how their comments changed other aspects of this paper, are presented later. Participants’ perspectives on the findings, together with sampling for maximum diversity of stakeholder perspectives, data triangulation and team involvement in analysis, added to the rigour of this work (Mays and Pope, 2000), as did our reflexive stance throughout (Symon and Cassell, 2004). Ethical approval for the SPIRIT trial and process evaluation was granted by the University of Western Sydney Human Research Ethics Committee, approval number H8970.

Results

While it is impossible to fully disentangle their impact from other contextual factors, it is evident that LPs made a profound difference to the way that SPIRIT was communicated, perceived and engaged with in each of the six intervention sites. We present the findings in relation to our research questions:

1. What were the professional characteristics of the LPs?
2. How did LPs perceive and promote SPIRIT?
3. To what extent did LPs operate as champions, opinion leaders and/or boundary spanners?
4. How can we explain the variation in how LPs perceived and promoted SPIRIT? Including (a) What was the role of organisational leaders? and (b) What was the role of organisational context?

Some details about LPs and the organisations in which they work have been altered to preserve anonymity. We have deliberately obfuscated any details that may reveal which LP was based in which organisation and, in the case of multiple LPs in the same agency, their position in the sequence and whether or not they were the primary LP.

What were the professional characteristics of the LPs?

In five agencies, the CEOs nominated the most senior member of staff with a research or evaluation role to act as the SPIRIT LP. In two cases, these staff passed on the function to a more junior member of their team. In the sixth agency, which had no dedicated research or evaluation position, the LP function was held by a senior executive. Consequently, there was substantial variation in the hierarchical position and role of LPs within their organisations.
The pivotal position of ‘liaison people’

**Seniority**

Contrary to our expectations, LPs with greater seniority did not always facilitate higher levels of participation. Reminders sent by the most senior LP often resulted in a spike in online survey completions compared to a nil effect from less senior LPs, but there was no consistent relationship between seniority and survey response across the agencies. Interviewees speculated that simply appointing a senior person to act as the LP “spoke volumes” about the “authenticity” of that agency’s commitment to SPIRIT. How this seniority was used, however, was equally important. Leveraging power may have had adverse impacts such as causing resentment from staff who were instructed to attend workshops. Conversely, seniority enabled LPs to make executive decisions, whereas junior LPs needed to elicit responses through bureaucratic chains of command, often negotiating new processes for dealing with the questions SPIRIT posed. This meant they often took longer to complete core tasks, but this did not result in lower participation rates overall.

**Organisational role**

Interviewees stated that where the LP had a research or evaluation position it bolstered their credibility as an appropriate facilitator for SPIRIT, but in some cases this association constrained how SPIRIT was perceived. For example, in an agency where the LP was an evaluation manager some staff assumed SPIRIT was an evaluation programme:

> who that person is affects what you think the presentation is going to be about [and its] relevance to your team… because she might have clearly explained [the study] but in your head it’s evaluation.

These interviewees speculated that their LP’s position increased the intervention’s credentials as an evaluation resource, but reduced the likelihood that staff who were not involved in evaluation would participate. In another agency, several participants expressed discomfort about the purpose of the relatively new organisational position that the LP occupied, and indicated that this led to negative connotations for SPIRIT.

**LP coverage and workload**

There was considerable variation in the turnover and coverage of the LP function during SPIRIT. Two agencies had a single LP for the entire study period, but with lengthy absences in one case. Two had a single LP during the intervention phase, but different LPs during data collection. In the remaining agencies, multiple people acted as the LP across both phases, including periods during the intervention in which there was no LP. This impeded the conduct of the outcome measures and the selection of, and arrangements for, intervention sessions. It also appeared to impact negatively on staff awareness of the intervention, with several interviewees in these agencies stating they did not recall any communication from their various LPs about SPIRIT. In most cases, staff turnover or restructuring caused the LP function to be transferred, but in one agency it occurred because two LPs found the demands too
onerous. As one manager explained, “It ended up being a bit more work than we’d anticipated… she literally just couldn’t manage it all”.

These demands were not fully knowable in advance, partly because this was a novel trial, but also because LPs’ workloads differed hugely depending on how they consulted about tailoring the intervention. The SPIRIT team tried to minimise the burden on LPs but found that they had underestimated time requirements in some cases.

How did LPs perceive and promote SPIRIT?

Liaison people across the six agencies had strikingly different conceptualisations of the LP function which, in turn, shaped how they approached the tasks. In one agency, SPIRIT was implemented following a major restructure. The LP speculated that her colleagues would confl ate her newly formed team with SPIRIT: “… my feeling is that people will tend to judge [us] by how useful they find SPIRIT, but also, maybe, to judge SPIRIT by whether they are embracing [us] or not”. Accepting this blurred line, she focused on adapting the intervention so it could be integrated into the team’s planned activities and directly support their professional development goals. The LP in another agency saw the function as an extension of his research governance position, so he used the experience to further develop cross-agency networks, convene research-orientated forums and increase essential skills in “translating and negotiating”. A third LP had operational oversight of the organisation and conceived her core task as managing the study’s demands: “getting it done efficiently”. She used her authority to act as a buffer between the trial and already overburdened staff, and to maximise measurement responses, “if I say ‘Do it’, people will do it”. Another LP was appointed to ‘fix’ SPIRIT after a previous LP had failed to engage staff. She conceptualised the work as a mobilisation exercise that depended on “getting buy-in”, so she prioritised interpersonal persuasion and advocacy. The LP in the fifth agency, who managed a research team, argued that the LP function was primarily administrative and questioned how appropriate it was for someone in his role. He focused on the core deliverables and minimised other tasks, whereas the LP in the sixth agency had an equivalent organisational role but saw the LP function as a “natural fit”. She conceptualised her task as maximising the value of SPIRIT which meant “generating belief” among managers so they would persuade their staff to participate, and devising mechanisms to “embed” the intervention’s ideas in organisational practice, “I tried to get something out of each [workshop] that would stay, would hang around for us”. Thus, in all cases, it seemed that the LP function and its core tasks were perceived in relation to the organisational position and professional responsibilities of the people who were assigned to act as LPs. As we show later, this was further shaped by the study’s perceived alignment with wider organisational goals.

To what extent did the LPs operate as champions, opinion leaders and/or boundary spanners?

Championing SPIRIT

Of the six LPs who were in place for the majority of the intervention phase, four were champions for SPIRIT and two were not. Non-champions did not: communicate a vision of what the intervention could achieve, demonstrate commitment or enthusiasm...
The pivotal position of ‘liaison people’

for it, or attempt to engage others in supporting it (Howell and Boies, 2004). Of the further 13 people who acted as LPs temporarily during the trial, we estimate that four displayed moderate championing, three were clearly not champions, and the other six (some of whom were in position for a matter of weeks) are unknown. We base these assessments on: (a) LPs’ statements in interviews and/or informal conversations about the value of SPIRIT and how they approached the tasks; (b) observable differences in LPs’ levels of enthusiasm and approach to tasks during intervention activities (for example, how they introduced workshops); (c) LPs’ conversations with SPIRIT team members; and (d) interviews with other agency staff. There was close agreement between LPs’ self-reported attitudes and behaviours in relation to the intervention and how other agency staff perceived their LP.

Unsurprisingly, LPs’ regard for SPIRIT appeared to correlate with their ‘championing’ and this, in turn, had an impact on how the intervention was structured, promoted, attended and perceived in each agency. In the two agencies where none of the LPs were champions, there was less consultation, the intervention was promoted inconsistently, aspects of the tailoring were less successful, and participants’ views of the intervention expressed in interviews and workshop feedback forms were more negative than in the other four agencies.

Relationship quality

Congruent with the literature, which asserts that the quality of relationships between champions and their colleagues is positively associated with influence (Howell and Higgins, 1990), many interviewees reported that positive regard for their LP encouraged their participation, “[she] is such a wonderful person that you kind of want to do it for her”. Strategies LPs used to encourage goodwill participation included dropping by colleagues’ desks to request they complete a survey, chatting in the kitchen about SPIRIT goals and, in one case, negotiating an explicit quid pro quo. Where interviewees reported instances of participating as a direct result of their LPs’ request, it was usually in the context of informal personal interaction (rather than emails or generalised comments in meetings). This suggests that friendly near-peer LPs may be best positioned to encourage participation (Rogers, 2003). However, as we point out later, leaders also played a vital role in this dynamic.

Selling SPIRIT

One of the most noticeable differences between champion and non-champion LPs was the extent to which they ‘marketed’ SPIRIT, that is, creatively harnessed organisational information channels and used rhetorical strategies to make the intervention and outcome measures more appealing. One LP did a mini presentation for staff who had missed the introductory session. He admitted to “embellishing the [audit feedback] a little bit” to create buy-in. Another explained, “I can talk it up in a way that sounds like it’s not a hassle and it’s interesting – and look, it is good stuff we’re getting out of this that will help you in your work”. This LP sought out and spoke personally to every member of staff nominated for each of the outcome measure points (about 25–30 people on six occasions), and achieved a 100% response rate.

One of the more senior LPs was also keen to increase response rates, so she sent a rare personal memo to staff telling them that, uncharacteristically, she had completed
the survey because the organisation needed good data for cross-agency comparison. In interview she explained her strategy: if staff knew she had completed a “bloody survey” they would understand its importance, plus they are highly motivated by competition. This agency’s response rate increased significantly and we are not aware of other factors that could explain it. This accords with findings that champions use formal and informal methods of communication to frame interventions strategically in terms of organisational orientations and objectives (Hendy and Barlow, 2012; Howell and Boies, 2004). Conversely, non–champion LPs may have undermined SPIRIT at times by overtly distancing themselves from the study. For example, one LP introduced a workshop saying that he didn’t know what it was about, and another forwarded email requests to colleagues about the online survey with the disclaimer, “Don’t shoot the messenger!” As we show later, these activities were influenced by LPs’ concerns that SPIRIT might damage their professional reputation.

The blurred distinction between persuasion and imposition was noted in every agency. All LPs admitted to “cracking the whip” to some degree, and most reported that staff sometimes felt hassled by multiple requests to participate. Overarchingly this related to “trying to get people involved with something that they don’t necessarily see benefits them directly”. Some LPs argued that getting colleagues to see these benefits placed too much responsibility on them; they asked, “Whose job is it?”. Naturally, LPs who valued SPIRIT were more willing to sell it and to convince colleagues to participate in data collection as part of a trade–off, but they also had more ammunition with which to do so. “Chasing” colleagues required “resilience”, but there is little doubt that these persuasive strategies increased participation; as one of the “chased” participants explained, “people find it hard to say no because… [the LP] is very politely persistent in that she’ll find you and hassle you until you [say yes]”.

These findings highlight the ethical challenges of workplace interventions. As others acknowledge, the line between persuasion and coercion is particularly delicate in organisational research where co–workers have recruitment responsibilities, protocols cannot be easily enforced by the research team, and where staff may regard participation as expected (Aguinis and Henle, 2004). SPIRIT sought to minimise coercion risks by reiterating the voluntary nature of participating and providing opt out opportunities pre– and post–data collection. Although managers could see who attended sessions they did not know whether their staff participated in data collection. There were no complaints about coercion, and the low survey response rates in most agencies suggest that people did not feel compelled to complete them, but in supporting the study a minority of LPs and managers may have strayed over this line, resulting in some unwilling participants (albeit in a study with negligible risks of harm).

Understanding SPIRIT

Being an LP was a learning curve and many of the longer serving LPs found that they became more adept over the duration of the study. Familiarity with the outcome measures increased the efficiency with which they were administered, just as experience of the workshop consultation and delivery process increased understanding of how components could be adapted. “[Before this] I couldn’t envisage what a Research Exchange would look like – what the possibilities were”.

Some non–champion LPs were unable to explain the study to their colleagues. One did not know what was happening in the intervention or measurements. Another
seemed unaware that his agency had choices about the intervention content: the selection of which he was meant to be facilitating. Many of the LPs found the study information dense and excessive so concluded that grasping it was not a worthwhile use of their limited time. One of them minimised the need to understand the study by telling her staff if they had any questions they should talk to the SPIRIT team. Another handed over the LP function to a colleague when action was required. This contrasts with one of the champion LPs who so delighted in knowing everything about the study that she playfully asked us to test her on the details.

**SPIRIT support for championing**

All but one LP described the SPIRIT team as supportive, but several felt the team could have done more to build relationships and anticipate their need for succinct, shareable information. Support was also not always consistent: some LPs who took on the function during the intervention period received less instruction than their predecessors. In one case, the LP felt this impeded her ability to champion SPIRIT:

> You should have really sat me down and said, okay, this is what it’s all about… That would have clarified the whole thing to me and I would have been able to say, okay, I can explain it to everybody and promote it, advocate for it, I suppose, which I don’t feel I have really been able to do.

LPs made suggestions for improving communications and support, highlighting the need for more on-site visits and face-to-face conversations, particularly in the early stages of the trial.

**Opinion leadership**

Participant interviewees described the characteristics by which they judged the suitability of their LPs; these spanned championing and opinion leadership and were contingent on two related concepts of legitimacy: credibility and commitment. Colleagues in one agency argued that it “didn’t make sense” for SPIRIT to be promoted by their LP, given her seemingly limited understanding of its aims and her indifferent attitude towards research in general. Conversely, colleagues of another LP commented that she was “ideal” given her “research credibility” both as an academic and an enthusiastic advocate for research-informed policy. The LP in a third agency concurred, explaining the CEO nominated him because he is publicly “committed to evidence” and known to influence colleagues’ engagement with research. All but one of the primary LPs thought LPs needed research experience in order to speak authoritatively about SPIRIT, and all regarded knowledge of the organisational culture as critical. There was no indication that LPs either were or were not viewed as opinion leaders in relation to other aspects of organisational business.

**Boundary spanning**

SPIRIT asked LPs to function as boundary spanners across different parts of their own organisations as well as across the agency–SPIRIT divide, hence LPs who were newly employed members of staff were at a disadvantage: they “couldn’t leverage
existing relationships” or make informed judgements about which colleagues and what documents were eligible for the outcome measures. Lack of familiarity with workplace culture and communication styles complicated consultations about how best to use SPIRIT but, as LPs’ relationships evolved, appeared to have less impact on later phases of the study. For example, a newly employed LP was keenly aware that she lacked essential workplace knowledge; yet, by the time of the post-intervention interviews, she was seen by colleagues as a highly effective networker and “ambassador” for SPIRIT. This LP had used SPIRIT to initiate organisational connections and had formalised boundary spanning by recruiting colleagues across the agency to act as team advocates for SPIRIT.

For an LP to bridge the divide between their organisation and SPIRIT, some “translation” was required. Most of the LPs attempted to make the study terminology and underpinning concepts more accessible, for example, they interjected during workshops to explain terms and provide illustrative examples. They also provided reassurance such as when, in the more clinically-orientated agencies, LPs used a drug trial analogy to illustrate that an intervention was being tested, not the participants.

**Mediation**

Although the SPIRIT team had mechanisms for communicating to agencies, they were dependent on LPs for conveying communication from agency staff. Consequently, lack of boundary spanning by LPs in some sites meant the SPIRIT team had no access to participants’ views and concerns (the process evaluation did not provide this feedback until after the intervention). Conversely, the more enthusiastic LPs acted as mediators, which increased the extent to which concerns were aired, addressed and fed back. For example, when the online survey was shortened one LP framed it as the researchers’ response to criticisms raised by agency staff. She informed her colleagues: “See, if you do have any questions or comments at any time about SPIRIT then you can tell me about them because they are listened to, and this is evidence of that”. It seems likely that these staff would have perceived such feedback as a validation of their participation. Staff in organisations with less communicative LPs might have welcomed the shorter survey but would have had less sense of agency in bringing it about.

**Brokerage and advice**

Variations in boundary spanning resulted in very different levels of advice from the LPs, and this impacted the SPIRIT team’s sensitivity to each organisation’s culture. Participants in one agency criticised SPIRIT for not using professional learning techniques that were their standard practice (they wanted direction via “pre-readings… so we can come into the room with our heads in the correct space”). These could have been incorporated if we had known. A forthcoming LP might have informed us about these norms unbidden, but we missed an opportunity to learn from the agency prior to the intervention about how to optimise activities in their setting. So while the trial benefited from boundary spanning LPs who proffered advice and creative suggestions, if we had acted as better boundary spanners ourselves we may have been able to tap into valuable insider knowledge more effectively across all the agencies.
SPIRIT team responsivity

Effective boundary spanning was a two-way street requiring mutual responsiveness and conciliation. LPs identified four behaviours from the SPIRIT team that they found particularly encouraging: 1) SPIRIT staff sending positive reinforcing feedback about the LPs’ hard work and positive impacts to their manager/CEO; 2) small appreciative gestures from the implementation team such as ‘thank you’ emails, verbal acknowledgements during workshops, and gifts of chocolates at Christmas; 3) changing aspects of the trial in response to agency feedback (for example, shortening the online survey); and 4) Supporting LPs to use their expertise to adapt information materials and participation strategies, “the good thing was that [the SPIRIT team] always acted on what I suggested… [they] realised that I know the organisation better than they would and what works here”. Thus the positive interactions between LPs and the SPIRIT team were co-adaptive. Where LPs’ suggestions were not acted on (usually due to infeasibility or adherence to the study protocol) this caused frustration. Clearer communication about why those decisions were made might have lessened this irritation and provided the LPs with a rationale they could share with colleagues.

How can we explain the variation in how LPs perceived and promoted SPIRIT?

Cost/benefit judgements

LPs made informal cost/benefit analyses about the potential value of the SPIRIT intervention for their organisation in relation to its demands. This determined their levels of enthusiasm for the intervention, how they perceived the LP function and how they approached its tasks. For example, one of the champion LPs was explaining her hope that SPIRIT would “pay off”: “… it’s certainly helped the general direction that we want to travel in terms of the role of research. So in that sense, yes. It’s been fairly time-consuming for me personally, but probably worth it for the organisation”. Perceived costs and benefits were influenced by management attitudes and behaviours, and by other organisational factors as described below, but were also entwined with an assessment of the potential professional benefits and risks in being associated with SPIRIT: those who expressed most enthusiasm about organisational benefits also identified value for themselves in being the LP. This assessment was particularly evident in two agencies in which the LPs were new employees. The one with a positive view of SPIRIT embraced the LP role, anticipating that it would help her develop internal connections and stakeholder relationships that were critical for her day-to-day work. After the intervention she reported that it had done just that. Whereas the other LP tried to minimise the risk that SPIRIT would be perceived as his project: “I didn’t want that connection”. As a new employee with no established organisational reputation it was uncomfortable to be associated with activities that he regarded as demanding with dubious merits. In three agencies LPs saw SPIRIT as a resource that bolstered their extant work in developing organisational research or evaluation capacity and, post intervention, they identified positive impacts in relation to their work. The least enthusiastic LPs did not identify positive impacts for their agencies or themselves. The ‘risk minimising’ LP described above said that the role had “helped expose me and connect me with people”, but not in the manner he would have chosen.
Being nominated

There was no association between how people came to be the LP and their attitude towards it. The only self-delegated LP was among the least enthusiastic. Conversely, the LP who was ‘volunteered’ in her absence went on to engage an overtly disengaged organisation and to facilitate one of the highest proportional attendance and survey response rates overall. Her initial view of SPIRIT as a confusing ‘research thing’ was far from enthusiastic, “[When] I got back from holidays and I was asked to take it over I was, like, ‘Oh my God! Why?’” Despite this inauspicious start, she strove to learn about SPIRIT and became convinced that her organisation could benefit. Motivated by this and the challenge of turning around the previous LPs’ lack of success, she approached the LP tasks with gusto and was able to incorporate a ‘conversion narrative’ as part of her rhetoric, “I’d say, ‘look I thought the same as you… what a hassle! But… it’s actually much easier than you think’.” This echoes findings that ‘change cynics’ who revise their views of an intervention can become highly effective champions (Hammond et al, 2011). It is also another ethical grey area in that several LPs were reluctant participants.

What was the role of organisational leaders?

Permission to push

Although LPs were asked to be the ‘face’ of SPIRIT in their agency, perceptions of the extent to which they were representing managerial views were key. Three LPs said they felt justified in being assertive about SPIRIT because it was known to be on behalf of the organisation’s leaders, “They knew it was something that I was pushing, but not for my own agenda… I was nagging them on behalf of our upper management”. In cases where leaders explicitly demonstrated support for the LP’s SPIRIT-related activities, they felt this ‘imprimatur’ was strengthened. Colleagues in these organisations concurred. According to interviewees across all agencies, the most persuasive incentive for completing the outcome measures was being asked by a well-liked, well-respected colleague who saw the endeavour as worthwhile, backed by evident managerial/CEO support.

In contrast, another LP expressed discomfort about the burden of repeated outcome measures and his need to cajole staff to complete them. Despite strong CEO espousal for SPIRIT, managerial support in general was not as visible or consistent as in some other agencies. For example, the LP’s immediate manager expressed scepticism about SPIRIT during workshops, which probably contributed to a less conducive environment for persuading colleagues to participate. As others have found, managerial cynicism can depress staff attitudes to organisational change initiatives (Rubin et al, 2009). Even where LPs perceived managers as committed to SPIRIT, they often struggled to get visible backup: “The main challenge for me, I think, is engaging our leaders enough so that they can convey the message to staff”. And in some cases, managers constrained LPs’ initiatives, limiting SPIRIT’s reach in the process. Examples included refusing an LP’s request to introduce a ‘SPIRIT slot’ at team meetings, and instructing the LP to reduce burden on the agency by limiting the number of staff who were invited to participate.
SPIRIT and work performance

The extent to which SPIRIT was formally recognised as part of the LPs’ work varied. In most cases managerial oversight of SPIRIT was added to the LPs’ usual reporting lines. In three agencies, LPs and their managers identified ways that SPIRIT could be used as an opportunity for professional development, for example, using LP activities as a vehicle for increasing their status and/or exposure in the organisation, and building SPIRIT deliverables into performance reviews. Such strategies strengthened these LPs’ desire to make it work. Where LPs radically shaped SPIRIT to address organisational priorities this was possible only because the LP already had some responsibility for developing such initiatives, and there was managerial support for using SPIRIT this way. Figuring out how to accommodate and use SPIRIT within LPs’ work was dependent on managers understanding the scope and responsibilities of the function and how these could be enacted in their organisational context, “It is one thing nominating a liaison person and then another thing to find, oh, does that liaison person have the authority to take decisions on all of these areas or to speak across the organisation? Or is their role more administrative?”

What was the role of organisational context?

Paradigmatic compatibility

Perceptions of SPIRIT’s compatibility with the organisation’s conceptualisation and use of evidence appeared to be the strongest determinant of why LPs saw greater or lesser value in SPIRIT. When interviewed, two of the most unenthusiastic LPs explained that the intervention made assumptions about how they should be engaging with research that did not align with their practice:

… people are operating at a different level from what is assumed [by SPIRIT], and have different needs. It’s no longer to do with access to research evidence, it’s what do you use and how do you use it to articulate good practice? How do you cut through the politics? How do you get people at the frontline to become aware of what they do and get them to throw back at you what kinds of questions are important, and how can that translate into research and policy? Which are very different kinds of questions from just how do you get more research into policy.

These LPs rejected the implication that their organisation should improve their use of research in the way SPIRIT conceptualised it, and did not believe that an externally designed intervention was an appropriate means of tackling highly situated knowledge-to-practice concerns. Their views were supported by other interviewees in the same agencies, suggesting that they were representative of their dominant workplace cultures. It is possible that, say, practitioners from other jurisdictions sharing real world experiences, or workshops that focused on internally developed research or evaluation, might have been more welcome. But the more disengaged LPs seemed unclear about how much intervention opportunities could be adapted and may not have considered these to be possibilities. In one case, the SPIRIT team pushed for
a workshop to be facilitated collaboratively with an expert in that agency. The idea was welcomed in principle, but later dismissed due to work pressures.

The more enthusiastic LPs worked in organisations that saw evidence, or the intervention, in a slightly different manner. Although all agencies had a pluralistic conceptualisation of evidence, an investment in stakeholder engagement, and extensive experience in implementing policies and programmes in messy real-world contexts, their emphases varied in accordance with their remit. Agencies working within specific biomedical fields (two of the intervention sites) seemed more disposed to embrace evidence-informed ideals than those with broad population health or systems reform briefs. This may reflect the extent to which forms of research often considered to be of highest academic quality – such as randomised controlled trials – could be applied instrumentally in their contexts. However, two of the champion LPs were in agencies with far broader remits. The first of these agencies was directly dependent on ministerial approval (and therefore, arguably, most susceptible to overt political pressure), yet their LP embraced SPIRIT. Several factors may have played a role. First, there were positive pre-existing relationships between the intervention designers and staff at different organisational levels who had commissioned some of the components offered by the intervention. Having used (and, to some extent, shaped) the product on offer, staff in this organisation were probably less likely to dismiss SPIRIT as pushing a purist and irrelevant evidence-based agenda. Second, the agency leaders enthusiastically and credibly espoused research utilisation and explicitly supported SPIRIT and the LP as a champion of the intervention. These factors were likely to reassure the LP that SPIRIT was sufficiently compatible with his agency to be worthwhile.

The second agency was embarking on training to strengthen their in-house research and evaluation capacity. The LP stated that their continued participation in the study was contingent on SPIRIT contributing to this pre-existing agenda, and she negotiated assertively to refashion intervention activities accordingly. Paradoxically, lack of established relationships between the SPIRIT team and agency staff may have facilitated this exchange as the agency had little to jeopardise in taking a strong stance. The commonality in all cases was the need for alignment between SPIRIT and the agency’s current engagement with research.

**Tailoring and alignment**

There was a strong sense of each agency being in flux and striving toward particular practice goals. This trajectory appeared to provide the benchmark against which LPs assessed the value of SPIRIT: given our circumstances and strategic goals, is this intervention worthwhile? To what extent does it provide opportunities that support how knowledge is conceptualised and situated within our day-to-day practice? This was echoed by interviewees’ predictions about whether they would participate in SPIRIT. To do so they would need to “see value”, gain “practical benefits”, and know that the intervention had “a direct relationship with the work that I’m doing”.

Programme flexibility and responsiveness was a key criterion for this assessment. The least enthusiastic LPs expressed concerns about structural inflexibility: “The tailoring of the programme is not really tailoring. What it is, you’re giving us a menu… you told us what you’re doing… and all we’re doing is ticking the boxes”. They saw limited scope for extensive adaptations because they regarded SPIRIT as fundamentally
non-consultative, “You’re talking to [us] but it’s a one way situation”. However, where LPs experienced the intervention as genuinely tailorable they maximised its benefits by working with SPIRIT staff to shape the workshops and resources to address organisational priorities. Two LPs integrated intervention components into a wider programme of staff capacity building, selecting topics, content and formats specifically to complement internal initiatives. Timing was also critical. Managers in the organisations in which these two LPs were based wanted the intervention to start at the same point as their internal initiatives, and one insisted on a hiatus while a major restructure was finalised. SPIRIT may have been better integrated by other agencies if organisational leaders could have decided when the intervention commenced.

**LPs’ view about our findings**

All six of the primary LPs who were invited to comment on the manuscript responded. Three gave general neutral or favourable feedback, and three commented more specifically. LPs were asked to alert us to any concerns about their identifiability but none did so (though one was initially concerned that other LPs might be identified). One LP asked for a word to be softened and another questioned an ambiguously phrased description of her agency. We agreed with their feedback and made amendments they were satisfied with. Two LPs developed themes in the manuscript about aspects of the trial that motivated them (belief in the goals of the trial, wanting to work with the trial leaders, leadership support and building LP tasks into their work performance review) and the characteristics required for the LP function (organisational and communicative skills, cross-agency connections and tenacity). This feedback has been included with the findings they relate to. Two felt that, having read what LPs in other agencies were doing, they would have benefited from interacting periodically during the trial to share experiences and discuss strategies: “after reading the manuscript, I really felt the loss of not having an opportunity to interact with other LPs – I think we could have learned a lot from each other!”

**Implications**

Our findings highlight some of the challenges of implementing complex interventions in real-world settings where the intervention’s ideas and activities must be carried by, and work through, existing organisational structures, processes and relationships (Damschroder et al, 2009; Locock et al, 2001). In such interventions change is a series of entangled interactions which are impossible to fully control (Greenhalgh et al, 2004). Nevertheless, it is also apparent that LPs wrestled with practical and, in some cases, conceptual obstacles which, in hindsight, could have been better anticipated by the intervention team. For example, we could not manage the frequency with which the LP function was transferred during the intervention, but we did know that policymakers change jobs rapidly and often act in other roles, so we could have designed a better system for supporting these transitions.

The conceptual obstacles suggest that more fundamental revisions should be considered. This will be addressed in further papers when all of our trial data can be considered. In the meantime, we acknowledge the dilemma that conceptual differences presented for some of the SPIRIT LPs. From one perspective, LPs’ views were self-fulfilling: where they judged the intervention to have potential value they invested
their efforts thereby adding value and experiencing SPIRIT as worthwhile. Where they judged it to have little value, little was added and little was experienced. But how much value could be added? The non-champion LPs were reflecting wider organisational concerns about the dissonance between an externally developed intervention that appeared to pre-frame the problem it was addressing, and in which experts provided generalisable knowledge when they saw knowledge as constituted through local practice: what Gabbay and le May (2011) call knowledge-in-practice-in-context.

It may be helpful to consider these findings in the light of previous research that shows even where change agents are highly respected opinion leaders their influence is bounded by current organisational norms and expectations (Rogers, 2003). It is far easier to motivate people who are receptive to the ideas presented in an intervention than those who are cynical (Rogers, 2003). Thus LPs may have been able to galvanise people’s engagement with SPIRIT positively or negatively but, without modifying the intervention substantially, could not have driven transformative change that countered dominant cultural tendencies, no matter how personally committed they were (Dibella, 2007; Hammond et al, 2011; Locock et al, 2001). The existing culture of research use within a policy organisation is known to affect how research utilisation intervention strategies are received. (Dobbins et al, 2009)

The findings from this study support those observed in other studies in that the delivery of interventions is profoundly affected by those who act in facilitation roles akin to that of our LPs (for example, Harvey et al, 2002; Ipsen et al, 2015; Kitson and Harvey, 2016). Further, that the LPs’ ability to function as champions, opinion leaders and boundary spanners, was critical. For example, Dixon-Woods and colleagues found that interventions were most effective when:

… those locally charged with implementation were sincere in their beliefs about the value of the program, were able to create transdisciplinary alliances, had local credibility among peers, were prepared to tolerate debate but exercise firmness, and used multiple tactics including role modelling, persuasion, sanctioning, reminders, and constant feedback. (Dixon-Woods et al, 2013)

What this study adds is an analysis of how these issues played out in a research utilisation trial in policy agencies. In particular, our findings suggest that concepts from political science about the contingent nature of evidence in policy (de Leeuw et al, 2014; Liverani et al, 2013; Pawson, 2006; Sanderson, 2009) and how it intersects with policy practices and organisational change (Armstrong et al, 2013; Evans et al, 2013; Hallsworth et al, 2011; Sundell et al, 2013) were central to how the intervention was facilitated. The paradigmatic compatibility of SPIRIT with agencies’ current and proposed research use strongly affected each LP’s views about the value of SPIRIT and this shaped how they engaged with and facilitated the intervention.

A standardised checklist of LP attributes is not meaningful in isolation, but we believe that some propositions (generalised theoretical statements grounded in the data (Ritchie and Lewis, 2003)) can be drawn from our findings. Given the complexities outlined above these propositions may be somewhat aspirational, but they point us in the right direction for identifying and supporting LPs in interventions similar to SPIRIT and, potentially, for informing a framework for evaluating attributes and
LP attributes

Proposition 1: The LP must believe that the intervention is worthwhile

At best, they will be genuinely enthusiastic about its merits – a champion. At least, they will judge that the benefits outweigh the demands.

As expected, the ideal internal facilitator for an intervention study such as SPIRIT appears to be a genuine champion (someone who believes in the intervention and will advocate for it energetically), an opinion leader (someone with informal organisational influence), and a boundary spanner (someone well-networked in their workplace who can also communicate effectively across the intervention–organisation divide). However, opinion leaders and/or boundary spanners who hold an indifferent or negative opinion of the intervention may undermine it (intentionally or unintentionally), while an enthusiastic champion is likely to ensure core tasks are delivered and amplify enthusiasm, albeit on a smaller scale than the opinion leader or boundary spanner. Consequently genuine support for the intervention appears to be a more important primary characteristic than influentially or connections. This is hard to ascertain up front and is dependent on local cost/benefit judgements, but managers are well placed to identify likely candidates and, in combination with intervention staff, encourage increased appreciation of the intervention’s potential. Others have reported success in gaining support from people who were initially opposed to an intervention (Hammond et al, 2011). Alternatively, agencies might issue an internal call for LP candidates, assuming that self-nominees are more likely to be committed to the intervention and the work required to facilitate it.

Proposition 2: The LP should have credibility in relation to intervention goals

Colleagues judged the suitability and effectiveness of their LP in relation to their credibility as an informed advocate for the intervention. Credible LPs had a professional reputation that aligned with the intervention goals (for example, they modelled and espoused research-informed work practices). This point and the one above accord with the literature which indicates that in order for a colleague’s espousal to be meaningful they must be perceived as someone who believes in what they are saying and knows what they are talking about (Dearing, 2009; Locock et al, 2001).

Proposition 3: The LP should have sound cross-organisational knowledge and connections

The intervention was more tailored, more creatively integrated, and better attuned to professional development expectations when LPs consulted with colleagues and shared their knowledge about organisational priorities, processes and learning norms with intervention designers. LPs’ ability to act as intervention intermediaries in this regard required them to have (or be able to rapidly acquire) a good understanding of their organisation and the people who work in it. This requires breadth: without boundary spanning skills, the efforts of champions may be restricted to highly localised contexts (Hendy and Barlow, 2012). But it also requires depth: an ability to understand diverse conditions. They are clustered in three categories: LP attributes, Managerial support and Intervention team responsibilities.
perspectives and needs arising from complex contextual interactions, and to respond accordingly (Harvey et al, 2002; Wilkinson and Frost, 2015).

**Proposition 4: The LP should have good interpersonal skills**

Ideally, they will be friendly, approachable and well-liked. Unsurprisingly, our data support assertions in the literature that people are more inclined to do things for people they like. This reminds us that ‘reach’ is about more than access. The quality of connections was just as important as the quantity for supporting organisational understanding and engagement, including identifying and resolving concerns during implementation. The need for communication and project management skills is a given.

**Managerial support**

**Proposition 5: Organisational leaders need to visibly back the LP as well as the intervention**

Strong, visible support for the intervention from managers was key in assuring LPs that their efforts – even when they verged on ‘nagging’ – were seen as reasonable and warranted (see also McCormack et al, 2013). Colleagues confirmed that strong support from above increased the LP’s authority and demonstrated they were acting on behalf of management. Others note that managerial support should encourage LP’s autonomy as overly specifying their approach could stifle enthusiasm and creativity (Markham, 1998).

**Proposition 6: If possible, the LP function should be incentivised within the organisation**

Enthusiasm for the intervention appeared to be enhanced when mechanisms or opportunities associated with the LP function benefited the LP professionally. This included formal professional development recognition (for example, building the work into performance indicators); increased organisational exposure, status or connections; or furthering the LP’s own work. In most cases, this will be effective only if there is some congruence between the intervention goals and the LP’s day-to-day work. Protected time for the LP tasks to be conducted during work hours should be agreed (Kirchner et al, 2012). A caveat: incentives should in no way pressure LPs to coerce participation.

**Intervention team responsibilities**

It is hard to overemphasise the importance of the relationships between the intervention team, the LP and organisational leaders. With the benefit of hindsight, these relationships would have been given a higher priority in our study.

**Proposition 7: Intervention staff should provide CEOs, LPs and the LPs’ line managers with clear and realistic guidance about the attributes and demands of the LP function**

The strikingly different conceptualisations of the LP role indicate that, at a minimum, we must emphasise that LPs are skilled facilitators rather than administrators *per se.*
Propositions 1–4 above provide the key messages for this exchange. The intervention team must describe the full scope of responsibilities and err in favour of over-estimating likely time commitments.

**Proposition 8: Agencies should be supported to enact the role of LP flexibly where it does not compromise implementation fidelity**

Our findings indicate the benefits of a flexible approach in which core objectives and tasks are specified but the strategies for achieving them can be developed locally (Haynes et al, 2016). For example, agencies might prefer to divide the LP function between two members of staff, with one taking responsibility for administrative tasks and another for creative input, persuasive communication, and higher level decisions. This has been effective in other studies, especially when those staff work (and are therefore likely to have influence) at different levels of the organisation (Kirchner et al, 2012).

**Proposition 9: Intervention staff should actively engage the LP in planning and problem-solving, treating them as a partner in the intervention rather than a conduit**

Where LPs shared detailed insider knowledge, employed creative strategies, and made suggestions for increasing the benefits of SPIRIT in their organisations, intervention activities were assessed by implementation staff and participants as more useful. This indicates that working with LPs as an intervention development partner, rather than as an implementer, would increase our ability to learn about and respond appropriately to local conditions, enhancing the relevance and fit of the intervention’s goals and activities (Howell and Boies, 2004). ‘Ownership’ approaches have been highly successful in effecting and sustaining change (Lopez-Patton et al, 2015; Zimmerman et al, 2013). They enable interventions to focus less on diffusing knowledge and more on contributing to how it is shaped and applied (Knights and Scarbrough, 2010). LPs who co-owned the intervention would be more likely to understand it fully (genuine dialogue bypasses much of the formal communication that SPIRIT struggled with), believe in its potential benefits and be perceived by colleagues as authentic advocates (Howell and Boies, 2004). However, this would depend on a fundamental philosophical agreement between the LP and the researchers about the goals of the intervention. Such collaboration would also require a significant time commitment.

In a subsequent paper, we plan to examine the concordance between the LP attributes, perceptions and behaviours reported here (findings which are blinded to the quantitative results) and the observed intervention effects. We recognise that many factors will affect how SPIRIT was received, but believe that the propositions outlined above, together with our analysis of the centrality of each organisation’s research culture and trajectory of change, will help explain the trial outcomes.

**Limitations**

A limitation regarding the interviews was our inability to reach everyone who acted as an LP in their agency, and to interview as many senior managers as we would have liked, including the agency CEOs. Hence we may have missed some important perspectives. We were also unable to test our propositions formally. Thus, while they
are sound representations of our findings across the six intervention sites, we do not know to what extent they provide useful applied guidance in identifying and working with LPs, nor how applicable they are to different organisational contexts.

Conclusions

This paper shows that the LPs who acted as facilitators of the SPIRIT study had a profound impact on how the intervention was implemented. LPs made informal cost/benefit analyses in which they weighed the value of participation against its demands and potential risks. Their different conclusions – influenced by their organisation’s mission, research utilisation norms, epistemological stance and leadership support – led to substantial variation in how they facilitated, promoted and tailored the intervention. This impacted on participation and engagement with the study across their respective organisations. LPs’ judgements about SPIRIT may have had a degree of self-fulfilment (they got what they put in), however, in some cases the intervention’s form and content may have been unsuitable for adaptations that could best address the organisations’ most pressing knowledge-to-practice needs. This indicates that the design of research utilisation interventions in policy agencies should incorporate potential participants’ views about the role of evidence in policymaking and how local practices can be best supported. Nine propositions were developed from the data that may assist in identifying and supporting facilitator roles in interventions similar to SPIRIT and, potentially, inform a framework for evaluating attributes and conditions.

Acknowledgements

Our sincere thanks to the busy staff in the six policy agencies that participated in SPIRIT and contributed to its process evaluation, particularly the liaison people. SPIRIT was funded as part of the Centre for Informing Policy in Health with Evidence from Research (CIPHER), an Australian National Health and Medical Research Council (NHMRC) Centre for Research Excellence (APP1001436) which is administered by the Sax Institute. CIPHER is a joint project of the Sax Institute; Australasian Cochrane Centre, Monash University; University of Newcastle; University of New South Wales; Research Unit for Research Utilisation, University of St Andrews and University of Edinburgh; Australian National University; and University of South Australia. The Sax Institute receives a grant from the NSW Ministry of Health. The Australasian Cochrane Centre is funded by the Australian Government through the NHMRC. AH is supported by an NHMRC Public Health and Health Services Postgraduate Research Scholarship (1093096).

References

Boeije, H, 2002, A purposeful approach to the constant comparative method in the analysis of qualitative interviews, Quality and Quantity 36, 4, 391–409
Caplan, N, 1979, 2–Communities theory and knowledge utilization, American Behavioral Scientist 22, 3, 459–70
CIPHER Investigators, 2014, Supporting policy in health with research: An intervention trial (SPIRIT – protocol for a stepped wedge trial, BMJ Open 4, 7
Conklin, J, Lusk, E, Harris, M, Stolee, P, 2013, Knowledge brokers in a knowledge network: The case of seniors health research transfer network knowledge brokers, Implementation Science 8, 7, 1–10
Damschroder, LJ, Aron, DC, Keith, RE, Kirsh, SR, Alexander, JA, Lowery, JC, 2009, Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science, Implementation Science 4, 50
De Leeuw, E, Clavier, C, Breton, E, 2014, Health policy – why research it and how: Health political science, Health Research Policy and Systems 12, 1, 55
Evans, BA, Snooks, H, Howson, H, Davies, M, 2013, How hard can it be to include research evidence and evaluation in local health policy implementation? Results from a mixed methods study, Implementation Science, 8
Grimshaw, JM, Eccles, MP, Greener, J, Maclellan, G, Ibbonson, T, Kahan, JP, Sullivan, F, 2006, Is the involvement of opinion leaders in the implementation of research findings a feasible strategy, Implementation Science 1, 3
Hammond, GD, Gresch, EB, Vitale, DC, 2011, Homegrown process improvement employing a change message model, Organizational Change Management 24, 4, 487–510
Haynes, A, Brennan, S, Carter, S, O’Connor, D, Huckel Schneider, C, Turner, T, Gallego, G, 2014, Protocol for the process evaluation of a complex intervention designed to increase the use of research in health policy and program organisations (the SPIRIT study), Implementation Science 9, 1, 1–12


Howell, JM, Boies, K, 2004, Champions of technological innovation: The influence of contextual knowledge, role orientation, idea generation, and idea promotion on champion emergence, *Leadership Quarterly* 15, 1, 123–43


Kitson, AL, Harvey, G, 2016, Methods to succeed in effective knowledge translation in clinical practice, *Nursing Scholarship* 48, 3, 294–302


Markham, SK, 1998, A longitudinal examination of how champions influence others to support their projects, *Product Innovation Management* 15, 6, 490–504


The pivotal position of ‘liaison people’

QSR International Pty Ltd, 2012, NVivo qualitative data analysis software: Version 10
Rogers, EM., 2003, Diffusion of innovations (5th edn), New York: Simon and Schuster
Rubin, RS., Dierdorff, EC., Bommer, WH., Baldwin, TT., 2009, Do leaders reap what they sow? Leader and employee outcomes of leader organizational cynicism about change, Leadership Quarterly 20, 5, 680–8
Sandelowski, M., 1993, Rigor or rigor mortis: The problem of rigor in qualitative research revisited, Advances in Nursing Science 16, 2, 1–8
Thompson, GN., Estabrooks, CA., Degner, LF., 2006, Clarifying the concepts in knowledge transfer: A literature review, Advanced Nursing 53, 6, 691–701
Williams, P., 2002, The competent boundary spanner, Public Administration 80, 1, 103–24
Williams, P., 2011, The life and times of the boundary spanner, Integrated Care 19, 3, 26–33
Appendix 13. Chapter Nine Manuscript

Policymakers’ experience of a capacity-building intervention designed to increase their use of research: a realist process evaluation
Background: An intervention’s success depends on how participants interact with it in local settings. Process evaluation examines these interactions, indicating why an intervention was or was not effective, and how it (and similar interventions) can be improved for better contextual fit. This is particularly important for innovative trials like Supporting Policy In health with Research: an Intervention Trial (SPIRIT), where causal mechanisms are poorly understood. SPIRIT was testing a multi-component intervention designed to increase the capacity of health policymakers to use research.

Methods: Our mixed-methods process evaluation sought to explain variation in observed process effects across the six agencies that participated in SPIRIT. Data collection included observations of intervention workshops (n = 59), purposively sampled interviews (n = 76) and participant feedback forms (n = 553). Using a realist approach, data was coded for context-mechanism-process effect configurations (retroductive analysis) by two authors.

Results: Intervention workshops were very well received. There was greater variation of views regarding other aspects of SPIRIT such as data collection, communication and the intervention’s overall value. We identified nine inter-related mechanisms that were crucial for engaging participants in these policy settings: (1) Accepting the premise (agreeing with the study’s assumptions); (2) Self-determination (participative choice); (3) The Value Proposition (seeing potential gain); (4) ‘Getting good stuff’ (identifying useful ideas, resources or connections); (5) Self-efficacy (believing ‘we can do this!’); (6) Respect (feeling that SPIRIT understands and values one’s work); (7) Confidence (believing in the study’s integrity and validity); (8) Persuasive leadership (authentic and compelling advocacy from leaders); and (9) Strategic insider facilitation (local translation and mediation). These findings were used to develop tentative explanatory propositions and to revise the programme theory.

Conclusion: This paper describes how SPIRIT functioned in six policy agencies, including why strategies that worked well in one site were less effective in others. Findings indicate a complex interaction between participants’ perception of the intervention, shifting contextual factors, and the form that the intervention took in each site. Our propositions provide transferable lessons about contextualised areas of strength and weakness that may be useful in the development and implementation of similar studies.

Keywords: Participant perspectives, Research utilisation, Process evaluation, Realist evaluation, Health policy
Background
This paper presents a realist analysis of how a novel, multi-component intervention trial designed to increase research use capacity, known as the Supporting Policy In health with Research: an Intervention Trial (SPIRIT), functioned in six health policy agencies. Data from a mixed-methods process evaluation is used to unpack the processes of engagement and participation that were hypothesised to mediate the intervention’s success. These intermediate impacts are conceptualised as process effects (see Box 1 for definitions). We do this by describing what was delivered in the intervention and what process effects were observed, then identify explanatory ‘Context + Mechanism → Process effect’ configurations that show how the intervention, and the trial more broadly, was perceived by participants, why this varied across the participating organisations, and how these perceptions affected receptivity to the intervention’s ideas and resources. A realist approach is used because it supports rigorous comparative analysis of how those targeted by an intervention make sense of what it offers, and how this is shaped by context [1–3].

**Box 1 Definitions of key concepts used in this paper**

| Context | In realist terms, context is any system, structure or condition that affects outcomes, including individuals’ attributes and social interactions [3] |
| Mechanism | Mechanisms are what makes an intervention work: “They are not the observable machinery of program activities, but the response that interaction with a program activity or resource triggers (or does not trigger) in the reasoning and behaviour of participants” [70] |
| Process effects | These are proximal impacts that influence intervention outcomes or are of evaluative interest for other reasons (e.g., they help explain unexpected variation in implementation); others use the term ‘formative outcomes’ [84]; Desired process effects are those the investigators consider to be prerequisites for a successful intervention |
| Programme theory | This is, “An explicit theory or model of how an intervention contributes to a set of specific outcomes through a series of intermediate results” [85]; programme theory should be plausible, useful and consistent with the evidence |
| Proposition | Propositions are generalised theoretical statements grounded in the data [86]; in realist evaluation, they link and condense information about contexts, mechanisms and outcomes; propositions are refined through empirical testing but remain fallible [87] |
| Realist process evaluation | Process evaluation helps explain how an intervention had its effects [7]; realist process evaluation applies realist principles to this process and investigates causal patterns (known as demi-regularities) to show how intervention strategies may be operating under what conditions to generate process effects for which groups [3] |
| Retroduction | This is a form of analysis that “involves constant shuttling between theory and empirical data, using both inductive and deductive reasoning” [88] |

Understanding interventions
Interventions – planned activities to change individual, group and/or organisational behaviour – are not passively received, but are actively shaped by the people who participate in them and the circumstances in which they are delivered [4–6]. Understanding the ways in which participants interact with and perceive an intervention is vital for determining how and why it was, or was not, effective [7]. This requires moving beyond measures of participant satisfaction – sometimes derided as “happy face evaluation” [8] – towards methods which delve into “the complexity, flux and contextual variation that inevitably occurs in real life situations” [9].

Many organisational capacity-building interventions fail because they do not take sufficient account of participants’ workplaces [10]. Successful interventions introduce strategies (ideas, activities and resources) that are contextually apt [7, 11] and which are therefore able to produce desired interactions [3]. For example, in organisational interventions, participants’ perceptions and interactions are affected by factors such as the organisation’s culture [12], its history of change [13, 14], staff heterogeneity [15] and trust in management [13].

Information about how implementation interacts with people and place over the course of an intervention is frequently overlooked [16]; yet, it is necessary for making informed assessments about the worth, adaptability and transferability of strategies designed to bring about individual or organisational change [9]. In multi-component interventions it is often impossible to disentangle which components were more or less effective, or what variations in combination might maximise effectiveness [17]. These interventions frequently trigger unanticipated causal processes and have unpredictable impacts that standardised measures are unlikely to capture [18]. This may be especially important for interventions where participants have involvement in the tailoring and/or delivery of an intervention, since their attitudes towards its content, form and goals are likely to have profound impacts on what is delivered and how it is received [19, 20]. Indeed, there is an established link between outcomes and the ways that participants gauge the quality of their involvement in tailoring the scope, content and process of flexible interventions [4].
[25], where an intervention promotes greater use of research, or claims to be evidence based, participants may actively critique that premise [26, 27]. Thus, determining if and how such an intervention is compatible with participants’ beliefs and practice norms is critical.

Despite these arguments, many interventions are reported (and, by implication, conducted) with minimal consideration of the interactions between the intervention activities, the people who took part, and the circumstances that mediated this relationship [9, 28]. As Clark et al. note, “Little research has explored individuals’ experiences of programmes or examined how programme dimensions lead to changes in behaviour...individuals’ meanings, experiences and reactions to the programme and the effects of their wider context are simultaneously disregarded” [29]. Realist process evaluation is well equipped to redress these oversights [1, 3].

The study being evaluated: SPIRIT

SPIRIT was a stepped wedge cluster randomised trial that tested the effects of a novel intervention designed to increase the capacity of health policy agencies to use research. Six organisations in Sydney, Australia, participated. Five were state government agencies and one was a national organisation funded by the federal government. An agency was eligible to participate if (1) a significant proportion of its work was in health policy or programme development, and (2) there were at least 20 staff involved in health policy, programme development or evaluation. A sampling frame was drawn from Government websites that listed all New South Wales and Australian government health policy and programme agencies located in Sydney. Members of the investigator team reduced this list to 16 potentially eligible agencies and ranked as highest those with the greatest focus on health and the largest numbers of relevant staff. The top six agencies were invited to take part, and all agreed [30]. Each agency’s Chief Executive Officer (CEO) signed an organisational-level agreement to participate in SPIRIT and nominated a liaison person: an internal member of staff who would be responsible for coordinating SPIRIT in their setting for the duration of the trial. There were six rounds of outcome data collection using three evaluation tools. These are described in detail elsewhere [30–35].

The intervention aimed to increase agency capacity to use research in relation to three goals, namely (1) the organisation and staff value research more; (2) more tools and systems are in place to support research engagement actions and the use of research; and (3) staff have greater knowledge and skill in research engagement actions and the use of research. SPIRIT’s design was informed by an action framework [36] and underpinning change principles that reflected composite theory from psychology, organisational science, adult learning and the research utilisation literature [30]. The intervention comprised multiple components hinging on interactive workshops such as research skills seminars, exchange forums with researchers, and a leadership programme targeting senior managers. Other activities included the provision of tools and resources (such as an online research portal); practice using systems for commissioning research reviews, analyses or evaluations; and CEO espousal of research-informed policymaking (Fig. 1). Agencies could choose options within and tailor many of the components to address local priorities. Each agency was asked to identify two lists of potential participants, namely (1) all staff involved in policy or programme development, implementation or evaluation who would be...
invited to take part in intervention activities and data collection and (2) managers who would take part in the leadership programme and promote SPIRIT.

An onsite introductory information session preceded the intervention and data collection in each site. The round of data collection that took place immediately before the intervention functioned as an audit and was followed by a feedback forum in which the lead investigator facilitated a deliberative dialogue with leaders about their agency's findings. Intervention goals targeting research engagement and use were identified during this process. Agency leaders considered how they would like to use SPIRIT's options to address these goals and, if applicable, any additional (non-SPIRIT) strategies for reaching their goals.

External research and policy experts were contracted to deliver workshops. They were briefed on SPIRIT’s 'change principles' and their workshop's objectives. The content of the tailored workshops was negotiated with the agency's liaison person, with input from presenters. Members of the SPIRIT research team coordinated the development and delivery of workshops and other intervention activities. Each site had a dedicated knowledge broker from the SPIRIT team who acted as the onsite 'face' of SPIRIT, negotiated tailoring and attended all intervention activities.

An in-depth, mixed methods process evaluation informed by realist thinking was conducted in parallel with the intervention. This paper is based on that data.

The role of process evaluation
Process evaluation investigates an intervention's implementation, change mechanisms and contextual interactions in order to explain (insofar as this is possible) how and why the intervention functioned as it did in each intervention site [18]. Process evaluation does not determine whether study outcomes are achieved, but it can identify process effects, namely proximal impacts of an intervention that make achieving outcomes more or less likely [37].

Aims
Using a realist evaluation approach [1, 3, 38, 39], we aimed to generate transferable learning in relation to the questions, (1) To what extent did SPIRIT achieve the desired process effects in each agency? and (2) How were these process effects generated? i.e. What mechanisms seem to account best for the patterns of engagement and participation observed across all agencies?

Methods
Realist evaluation
The SPIRIT process evaluation comprised a fidelity assessment and a theory-driven exploration of the interaction between the intervention, participants and the implementation circumstances, with the expectation that this would probably take a different form in each of the six agencies [40]. Theory-driven evaluation seeks to uncover causal pathways [41] and is well suited for understanding how multicomponent interventions function in complex real-world settings [42]. In this study, we adopt a particular theory-driven approach – a realist evaluation [43] – following the methods associated with Pawson [1], Pawson and Tilley [3], and others in the RAMESES II project [39]. Realist evaluation focuses on an intervention's underlying theory as its unit of analysis [1, 3], with the aim of determining "what works, for whom, in what circumstances, and how" [3, 44]. Realists posit that interventions introduce ideas and opportunities that generate effects in conjunction with participants' reasoning and resources. Thus, the interaction between intervention activities and the contexts of each intervention site will determine what (if any) mechanisms are activated and what outcomes (intended and unintended) are generated [45, 46].

We used a realist approach because it maximises the transferability of findings and operational learning from one setting to another (an enduring concern in intervention evaluation [47]), while also recognising complexity and the need to look beyond one-size-fits-all ways of responding to problems [1, 3, 48, 49]. Realist evaluation has been used effectively in studies of policy processes [50], implementation research [51], knowledge exchange [52] and evaluations of flexible intervention trials [19, 29], making it especially suitable for addressing the methodological challenges presented by a multi-component, novel and theoretically eclectic trial like SPIRIT (outlined in detail elsewhere [53]).

Importantly, analyses arising from realist evaluations are tentative, claiming only to be an informed hypothesis of "how something might be" [54] rather than a definitive version of reality. These hypotheses accrue plausibility when tested in further studies, but remain open to revision or rejection if alternative theories are more convincing [45]. In our study, data collection, management and analysis were concurrent; thus, we were continually testing and revising hypotheses within and across the six intervention sites over the 30-month study, but our findings are embryonic in realist terms.

Initial programme theory
Realist evaluation develops, tests and refines programme theory. SPIRIT was informed by a mixture of formal theory and experiential knowledge [30], and had both a well-articulated action framework [36] and clear principles about what should be provided [53], but did not offer hypotheses about the mechanisms that would generate increased capacity to use research. Based on existing trial materials and discussions with the designers, we articulated the overarching programme theory to make the intended causal pathway more explicit so that we could critique the assumptions underpinning the intervention design [1, 3, 55]. This was refined and agreed through further consultation:
SPIRIT will engage and motivate agency leaders to ‘own’ the intervention using audit feedback, deliberative goal-setting and program tailoring. This agency-driven approach will generate a priority-focused program that offers locally relevant practice support and accommodates differences in agencies’ values, goals, resources and remits. The program will comprise a suite of andragogical activities, tools, and connection across the research-policy divide that provide resources and build knowledge, skills and relationships. It will be supported via modelling and opinion leadership by agency leaders and dynamic external experts. CEOs will promote SPIRIT in their agencies and liaison people will facilitate the tailoring and implementation. These strategies will act synergistically to stimulate and resource participants at different organisational levels, leading to changes in values, practice behaviours and agency processes. This will facilitate increased use of research in policy processes.

This pathway informed the data collection, providing pointers about what to look for, but was used flexibly (rather than as a rigid investigative framework) as befits an exploratory study. We also looked for unintended effects, and considered alternative causal pathways that might better explain observed effects. The data offered the opportunity to develop a much richer understanding of the social processes and interactions than had previously been possible.

Process effects
The programme theory was used to identify desired process effects via discussion with the study designers. We then explored how these process effects were achieved in each setting for the range of targeted participants, or why they were not. Our conceptual framework for this work was informed by the implementation science literature that focuses on social processes and interaction in interventions (e.g. [6, 26, 56–60]).

Data collection
Causation, and the mechanisms that generate it, are seldom observable [3]. Therefore, in realist evaluation, data is triangulated to identify the interactive patterns that can most plausibly explain how the intervention led to the observed outcomes [61]. Quantitative data is helpful for identifying outcomes [1], while qualitative methods are usually necessary “to discover actors’ reasoning and circumstances in specific contexts” [62]. We used the following methods to capture information:

- Semi-structured interviews with 5–9 participants from each agency early in the intervention period (n = 33) and post-intervention (n = 43). Interviewees were purposively selected for maximum variation in work roles, attitudes to research and experiences of SPIRIT in order to explore the breadth of dimensions expected to influence interactions with the intervention [7]. Open-ended questions and prompts explored interviewees’ work practices and contexts, and their experiences and perceptions of SPIRIT, including their explanations for any change. The interview questions are available elsewhere [40]. This combination of context-, causal- and impact-focused questions across diverse participants was used to refine theory about what was working (or not), for whom and in what circumstances.

- Observations of intervention workshops (n = 59), and informal opportunistic conversations with participants before and after workshops. Workshops were audio recorded and field notes were written immediately afterwards. A checklist was used for fidelity coding through which we monitored the extent to which ‘essential elements’ of the intervention were delivered (detailed elsewhere [59]).

- Anonymous participant feedback forms (n = 553). These comprised Yes/No ratings on six statements: (1) The workshop was interesting, (2) The workshop was relevant to my work, (3) The workshop was realistic about the challenges and constraints of our work, (4) The presenter had appropriate knowledge and skills, (5) It is likely that I will use information from this workshop in my work, (6) It is likely that SPIRIT will benefit my agency (Additional file 1).

Some workshops had additional items, e.g. the forms for audit feedback forums included items about the clarity of the data and participants’ confidence that SPIRIT would be adequately tailored for their agency. All forms contained three open-ended questions: (1) ‘What worked well?’, (2) ‘What could be improved?’ and (3) ‘Any other comments?’ Forms were distributed prior to intervention workshops and completed immediately afterwards.

- Formal and informal interviews with the people implementing SPIRIT and the commissioned presenters.

- Limited access to information from the interviews conducted as part of SPIRIT’s outcome evaluation. These interviews focused on (1) organisational support for research use (n = 6), and (2) the role of research in the development of a recent policy or programme (n = 24). We reviewed transcripts from the first round of interviews (prior to the intervention), but thereafter were blinded to this data so that it would not influence the ongoing process evaluation analysis.

Data management and analysis
Qualitative data
Data was initially analysed for the whole process evaluation. Interview data was managed using framework
analysis [63] within NVivo v.10 [64] and used to develop descriptive case studies [65] in combination with data from the fidelity assessment, running memos for each agency, interviewee memos, the thematically coded data from field notes and the open-ended questions in feedback forms. These case studies described (1) each agency's context, change trajectory, workforce and practice norms, (2) their research use practices and culture, (3) how SPIRIT was implemented in each setting, and (4) the interactions between (1), (2) and (3). Framework categories and the structure of the case studies were iteratively developed from a priori concerns (such as the constructs the intervention was targeting and the hypothesised causal pathway), and from themes identified using inductive analysis [66, 67]. The method of constant comparison [68] was used to query and refine the initial programme theory and other emergent hypotheses throughout the trial. This work is described in more detail elsewhere [40].

Quantitative data
For each agency, we calculated the number and percentage of feedback forms responding ‘Yes’ to each of the six statements outlined earlier. In calculating these frequencies, the four different types of workshops (symposia, research exchanges, leaders’ forums and audit feedback forums) were aggregated.

Realist analysis
Using the data described above, we sought to explore the hypothesised pathway identified in the initial programme theory and to identify any other pathways leading to the interventions’ observed process effects, plus other impacts reported by participants or members of the implementation team [42].

We employed a retrodictive analytical approach that attempts to explain phenomena by theorising about what mechanisms are capable of producing them [69]. This involves studying events "with respect to what may have, must have, or could have caused them. In short it means asking why events have happened in the way they did" [51]. In accordance with realist evaluation principles, we focused on the interaction of SPIRIT with features of each agency’s context that appeared most likely to have influenced process effects [42, 70]. We developed explanatory configurations of the patterns we saw in the data. In realist evaluation, these are typically called Context + Mechanism → Outcome configurations [1, 3], but because the ‘outcomes’ of interest in process evaluation are process effects rather than study outcomes, we have called them Context + Mechanism → Process effect configurations herein. Propositions were then developed to summarise each configuration. This work depended on using each type of data to query, explain and balance the other to reach as comprehensive as possible accounts of what happened and why [71, 72]. Original data sources were revisited as required.

These process effects were identified prior to the development of Context + Mechanism → Process effect configurations and were used as a starting point in much of the analysis – although realist evaluation depicts outcomes (or, in our case, process effects) as the final step in the sequence, the analysis tends to start by identifying effects, then working backwards to investigate the conditions (context and mechanisms) that caused them [73]. We traced connections to and from observed process effects asking ‘What caused this?’ ‘Why didn’t this unfold as anticipated?’ and ‘What best explains these different responses between agencies?’ Analysis involved looking for data that might indicate the absence or weak functioning of mechanisms as well as the presence of a mechanism. This was aided by Dalkin et al.’s [46] assertion that mechanisms may vary in intensity rather than simply being present or absent.

AH, who led the process evaluation, reviewed and coded all data sources. SB, who contributed to the process evaluation design and analysis throughout the trial, independently reviewed a proportion of interview transcripts and cross-agency fieldwork memos. Their preliminary Context + Mechanism → Process effect configurations overlapped extensively and were workshopped with further reference to the wider data set to develop agreed configurations. Further discussion with our co-authors resolved differences and refined the final findings.

This analysis relied on abductive reasoning [74], which is an iterative cycling between data and likely explanations that incorporates inductive and deductive processes. We looked for evidence of factual causal mechanisms, and for evidence that supported, discounted or nuanced current causal hypotheses both in real time (as the intervention unfolded) and retrospectively (reviewing data already collected). Throughout this process, we sought to identify where our evolving Context + Mechanism → Process effect configurations aligned with existing theory; we revisited the theories used to inform the development of SPIRIT, asking to what extent did these theories support the patterns we were observing in the data, and also considered other theories that might better explain our findings. See Additional file 2 for an overview.

Results
In this section, we describe the implementation of the SPIRIT intervention, outline the observed process effects, and then attempt to explain how these effects were generated using Context + Mechanism → Process effect configurations. Finally, we present the revised programme theory.

Implementation
As Additional file 3 shows, some aspects of SPIRIT were delivered with a high degree of implementation fidelity;
Table 1 Overview of SPIRIT’s process effects and data sources

<table>
<thead>
<tr>
<th>Desired process effects for the trial</th>
<th>Observed process effects</th>
<th>Supporting data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Leaders espouse SPIRIT and its goals</td>
<td>All CEOs disseminated initial information about their agency’s participation in SPIRIT, but only four had a continuing visible role in supporting the intervention, e.g. sending updates and attending workshops; some executive members participated in each site, but to very different extents ranging from a half hour ‘drop in’ to repeated and enthusiastic participation; many managers talked about SPIRIT in team meetings and encouraged their staff to attend</td>
<td>Interviews at two time points (early-intervention ‘context’ and post-intervention ‘perceptions and impact’), ad hoc conversations with participants</td>
</tr>
<tr>
<td>2. Liaison people facilitate the intervention effectively</td>
<td>The use of a liaison person was very effective in the sites where the liaison person was enthusiastic about SPIRIT; four of the six worked hard to promote, tailor and administer the intervention, harnessing insider knowledge and using creative strategies, whereas the other two did not tailor or promote the intervention as thoroughly and expressed negative views to colleagues about SPIRIT</td>
<td>Observations of workshops, interviews and conversations as above, feedback from the SPIRIT team about their communications with liaison people</td>
</tr>
<tr>
<td>3. Targeted policymakers participate in, and are receptive to, intervention activities</td>
<td>Participation levels were good in that they met the SPIRIT team’s expectations for each site; each agency targeted different groups for different components so proportions and types of participants varied, but liaison people were satisfied with attendance and were occasionally surprised by very high numbers; attendance at workshops averaged between 11 and 20 participants per workshop, with between 102 and 158 total occasions of attendance across the six sites; there was full participation in other activities (e.g. trialling the commissioned research services); receptivity varied tremendously within, but especially between, agencies; see next section for more details, including possible reasons</td>
<td>Quantitative fidelity data from observations (using check lists and sign-in sheets), observations, interviews and conversations as above</td>
</tr>
<tr>
<td>4. Participants actively contribute to the content of those activities</td>
<td>Where there was opportunity, participants contributed greatly to workshop content via questions, discussion and case examples; interactivity was limited on some occasions in all agencies, usually because the presenter provided few opportunities; in larger groups, more senior staff tended to dominate, but other participants said this was still useful. Some liaison people helped craft workshop content and provided agency-based case examples; one agency co-presented a workshop; the agency staff nominated to test the research commissioning service were actively involved</td>
<td>Observations of workshops, including descriptive accounts of interactions and dynamics</td>
</tr>
<tr>
<td>5. Participants identify potentially useful ideas, techniques and/or resources</td>
<td>94% of those who completed a feedback form said they found workshops to be both relevant to their work and realistic about policy challenges and constraints; many interviewees identified specific benefits from SPIRIT, including improved awareness of useful researchers and research resources, understanding of the evidence relating to a policy problem and access to existing agency resources</td>
<td>Participant feedback forms, observations of workshops, interviews and ad hoc conversations with participants and liaison people</td>
</tr>
<tr>
<td>6. Participants use, or plan to use, these ideas, techniques and/or resources</td>
<td>Workshops facilitated less discussion than intended about how learning might be applied, but 95% of participants who completed a feedback form agreed, “It is likely that I will use information from this workshop in my work”; some interviewees said they planned to use ideas or resources, and a few had done so, especially newer staff; three liaison people had managerial-approved plans underway for research-focused education and/or systems improvement, e.g. mandated consideration of research in policy proposals; two agencies had plans to use their commissioned research products</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Desired process effects for the evaluation</th>
<th>Observed process effects</th>
<th>Supporting data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. Liaison people facilitate data collection effectively</td>
<td>All liaison people facilitated data collection sufficiently, although it was occasionally delayed and required prompting where liaison people championed SPIRIT they used additional strategies to encourage participation in data collection, in one agency this achieved a 100% response rate</td>
<td>Outcome measures completion figures, interviews with participants and liaison people, feedback from SPIRIT team</td>
</tr>
<tr>
<td>8. Targeted participants take part in data collection</td>
<td>In all agencies, there was full participation in the two interview-based measures, but more variable responses to the anonymous online survey; response rates dipped in the second measurement point, but stabilised after the survey was shortened; overall, the online survey response rate was 56% and there was a mean 74% response rate for process evaluation feedback forms; only three-quarters of invitees took part in a process evaluation interview</td>
<td>Outcome measures completion figures, interviews with participants and liaison people</td>
</tr>
<tr>
<td>9. The benefits of the intervention are judged to outweigh the burdens of the trial</td>
<td>Interviewees differed considerably in their assessments of the intervention, but where they felt it had benefits these were deemed to outweigh the trial’s burdens, this included those liaison people who championed SPIRIT from the start; workshops with high profile and dynamic ‘service-orientated’ presenters were especially valued; nearly 98% of all feedback form respondents agreed with the statement, “It is likely that SPIRIT will benefit my agency”</td>
<td>Early-intervention and post-intervention interviews, ad hoc conversations with participants and liaison people, feedback form data</td>
</tr>
</tbody>
</table>
indeed, every agency received audit feedback and the intended number of components on the topics they requested. Intra-organisational processes that were outside the control of the implementation team had greater variation. The promotion of SPIRIT and much of its administration depended on the attitudes and behaviours of liaison people and each organisation’s leaders, and to a lesser extent, the expert presenters commissioned for each workshop. This resulted in some loss of SPIRIT’s theoretical fidelity, i.e. the extent to which the intervention delivered its ‘essential elements’ (these are discussed in more detail elsewhere [53]). For example, the essential elements stipulated that workshops should be non-didactic and therefore the presenters should encourage participants to contribute as much as possible. Many workshops were highly interactive, such as the deliberative audit feedback forums, but others were not. This was because (1) the expert presenters sometimes overrode their briefing to facilitate discussion; (2) liaison people occasionally tried to maximise value by cramming content into workshops, which limited opportunities for participation; and (3) unexpectedly, the agencies seldom took up offers to co-design and co-present workshops.

In some sites, SPIRIT’s reach was constrained more than anticipated. Agency 6, for example, chose to focus some components of the intervention on one group of staff and limited participation accordingly. In Agency 3, managers attempted to minimise the onerousness of data collection by excluding some eligible staff from invitations to complete surveys. Agencies also defined their leadership groups quite differently, resulting in wide variation in the numbers and organisational roles of participants in the leaders’ programme.

**Process effects**

Table 1 describes SPIRIT’s process effects, i.e. the actions, behaviours and responses hypothesised to be necessary for SPIRIT to generate the capacity-related outcomes measured in the trial. Column 1 lists the process effects both for the intervention and the trial evaluation; we include the latter because of their impact on the quality of the evaluation and the way that SPIRIT as a whole was perceived. Column 2 presents a summary of our observations about the extent to which these process effects occurred. Column 3 shows the data sources for our observations.

**How were these process effects generated?**

We identified nine primary causal mechanisms (Fig. 2). The Context + Mechanism → Process effect configurations for each mechanism are presented in the following section. Each of the configurations begins with an overview of the context pertaining to that mechanism, a description of how we believe the mechanism functioned, how it generated process effects and how process effects differed between participating agencies. A proposition that summarises the hypothesised casual pathway precedes each configuration.

Cross-references to other mechanisms are in shorthand so that mechanism 1 reads as M1, etc. Similarly, agency numbers are shortened so that Agency 1 is shown as A1, and so on. Inevitably, this is a highly truncated presentation of our findings. For those who seek more

---

**Fig. 2** Overview of context-mechanism-process effects in the SPIRIT trial
detail, a narrative description of the data that informed our identification of each mechanism can be found in Additional file 2. This additional file provides an ‘evidence link’ between the data and the findings that follow.

**Mechanism 1**
Accepting the premise (Table 2)

**Mechanism 2**
Self-determination (Table 3)

**Mechanism 3**
The value proposition (Table 4)

**Mechanism 4**
“Getting good stuff” (Table 5)

**Mechanism 5**
Self-efficacy (Table 6)

**Mechanism 6**
Respect (Table 7)

**Mechanism 7**
Confidence (Table 8)

**Mechanism 8**
Persuasive leadership (Table 9)

**Mechanism 9**
Strategic insider facilitation (Table 10)

### Table 2 Mechanism 1 - Accepting the premise

<table>
<thead>
<tr>
<th>Context</th>
<th>How did mechanism 1 function?</th>
<th>How did mechanism 1 generate process effects?</th>
<th>Agency comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Each agency had existing goals, values, resources, practices and change trajectories. They viewed their capacity to use research, and the importance of increasing this capacity, quite differently. There were diverse norms about what evidence is and how it should be developed, which were affected by their primary stakeholder groups. In the wider environment there was increasing emphasis on action-based research (partnering with practitioners to produce research fit for immediate decision-making).</td>
<td>Potential participants accepted or rejected the premise of SPIRIT based on: 1. The compatibility of SPIRIT’s goals &amp; assumptions with local conceptualisations of evidence and its role in policymaking, including how research related to the agency’s remit, values and practice norms; stakeholder relationships; and change trajectory. 2. The compatibility of SPIRIT’s form – whether its design was congruent with local conceptualisations of ‘good’ or ‘appropriate’ intervention/trial models. 3. Relative advantage: if participants believed they or their agency would benefit from increased use of research, i.e. they saw a need for SPIRIT.</td>
<td>When they accepted the premise of SPIRIT leaders were more inclined to espouse the intervention and its goals. Liaison people facilitated the intervention more effectively, and targeted policymakers were more enthusiastic about participation and receptive to content. Where policymakers dismissed SPIRIT’s premise they said they were unenthusiastic about participation (yet many did participate) and had low expectations of content. Liaison people who rejected the premise admitted they did not champion SPIRIT or facilitate the intervention as effectively as they could have, but they managed data collection satisfactorily. Leaders who had reservations about the premise tended to express their doubts to staff, but also encouraged staff to participate in specific intervention activities.</td>
<td>Most potential participants in A1, A5 and A6 saw SPIRIT as addressing “a real need” and were open to what it had to offer. A3 staff supported the premise but many felt it did not apply to them as they had “no room for improvement”. A few interviewees in all agencies dismissed the premise of SPIRIT, but especially in A2 and A4. However, many of these participated in at least one workshop, either because it was expected or because the potential merits of individual workshops (M3) overcame reservations about the intervention/trial as a whole.</td>
</tr>
</tbody>
</table>

As others have noted, separating interactive processes into discrete mechanisms, while useful for theory building, fails to reflect their interdependence [61]. Many of the nine mechanisms include related concepts, which in some cases may be nested. For example, ‘self-determination’ (M2) is linked with ‘respect’ (M6) and may function as a mechanism within ‘self-efficacy’ (M5). Figure 2 illustrates feedback within our model. This accords with the realist view that contexts, mechanisms and outcomes are not fixed entities but are contingent on the focus of the current evaluation, i.e. they function as a context, mechanism or outcome in a particular part of the analysis. Thus, many of our process effects feed back into and overlap functionally with the identified mechanisms, and may well function as mechanisms when this data is combined with the study outcomes. This is especially pertinent in a process evaluation given that process effects are hypothesised to mediate the intervention outcomes. An example of feedback is our finding that ‘persuasive leadership’ is a mechanism, despite one of the process effects being ‘Leaders support SPIRIT’. This is because we found ‘persuasive leadership’ to be crucial in activating other mechanisms (e.g. in asserting SPIRIT’s value proposition) and thus in achieving many of the other process effects.

### Table 3 Mechanism 2 – Self-determination

<table>
<thead>
<tr>
<th>Context</th>
<th>How did mechanism 2 function?</th>
<th>How did mechanism 2 generate process effects?</th>
<th>Agency comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Externally designed interventions often feel imposed. SPIRIT aimed to enable agencies to identify local goals and tailor workshop content, but agencies were time pressured. Participants had extensive expertise in crafting policy, and many were experienced program designers and implementers using increasingly collaborative, bottom-up models. They viewed SPIRIT through this professional lens. The intervention’s start date was randomised. CEOs decided if their agency would participate and nominated liaison people. Agencies encouraged participation but, as others note, participation in organisational interventions can feel ‘expected’.</td>
<td>Self-determination is the feeling of having control 5. Interventions that foster self-determination share power and allow participants to pursue a variety of goals. Participants felt they had some control in relation to SPIRIT when there was: 1. Flexibility – scope to use the intervention to address their needs. 2. Decision-making support – audit feedback and deliberative processes helped leaders to make informed decisions, and those involved in tailoring received guidance about what could be achieved and how to do it. 3. Locally shaped content – managers and liaison people actively tailored goals and content, and ensured colleagues had a say in it (M9); interactive workshops enabled participants to drive content; and participants co-designed and co-presented workshops. 4. Choice about whether to take part in the intervention and data collection or be a liaison person, irrespective of managerial expectations. Self-determination had to be tempered with judicious decision-making. Some choices backfired such as when liaison people crammed content into workshops which overwhelmed participants.</td>
<td>Tailoring and interactivity were consistently viewed as critical for getting value out of participation. Self-determination encouraged leaders and liaison people to champion SPIRIT, and tailoring gave them a key selling point – “It’s designed for us”. When self-determination was constrained (e.g. in didactic workshops) it tended to frustrate participants. Tailoring was time-consuming and sometimes required new decision-making pathways, so some agencies found it burdensome and did not make full use of the flexibility on offer (even though they were adamantly it was necessary). Where liaison people rejected the premise of SPIRIT (M1) or believed that it was not flexible enough, they put less effort into tailoring and promotion. Leaders wanted choice about when SPIRIT started so it could be used more strategically to complement (or avoid) other activities.</td>
<td>Agencies 1, 3, 5 and 6 felt they had enough scope to shape intervention content and did so moderately to extensively. Where there was scope, participants in all agencies contributed to workshop content, increasing its relevance and applicability. Greater involvement seemed to increase receptivity and investment in outcomes. Less effort was made to tailor the intervention in A2 and A4. This may have increased expectations of (and actual) incompatibility. Several liaison people were reluctant to take on their role, and many participants in A2 felt obliged to participate. Despite considerable encouragement, only A5 co-presented a workshop. Two agencies modified non-flexible aspects of SPIRIT: A3 insisted on different participant eligibility criteria to reduce the burden on staff, and A6 requested a hiatus in the intervention while they managed a restructure.</td>
</tr>
</tbody>
</table>


We also concluded that mechanisms functioned on a continuum that encompassed negative and positive expressions. Mechanisms were activated to different extents in each agency and, on occasion, were activated negatively. For example, several interviewees made it clear that mechanisms such as ‘Self-determination,’ ‘Getting good stuff’ and ‘Respect’ were activated negatively when they were instructed by their manager to attend a 2-hour workshop that had no relevance to their work.

Revised programme theory
These results enabled us to revise our programme theory to reflect contextual contingency, which also increases the operational transferability to other interventions and settings (Table 11).

Discussion
From the participants’ perspective, the most positive attributes of the intervention were useful (i.e. relevant and applicable) content, high profile experts who delivered pragmatic content and demonstrably “got it”, active participation in intervention activities, and intervention flexibility supported by deliberative audit and feedback that informed goal-setting and customisation. Much of SPIRIT’s implementation fidelity was sound – all the components of the intervention were

---

Table 4  Mechanism 3 – The value proposition

<table>
<thead>
<tr>
<th>Context</th>
<th>How did mechanism 3 function?</th>
<th>How did mechanism 3 generate process effects?</th>
<th>Agency comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>The ideal of research-informed policy was espoused in the wider environment, but each agency had a distinct organisational culture that interpreted this differently. Busy policy staff were juggling competing demands and needed a good reason to take part in non-essential activities. They calculated trade-offs: ‘what can I afford to lose or postpone to make way for SPIRIT?’ All suffered generalised information overload, but many complained about the lack of useful research in their area. The challenge of explaining SPIRIT was exacerbated by: 1. a complex and unfamiliar study design; 2. flexibility (it was being tailored and in flux); and 3. two levels of outcomes: those of the trial (fixed) and those identified by the agency (targeting local goals).</td>
<td>A “value proposition” (promised advantage) is a convincing argument about the worth of a strategy that is assessed by prospective users on the basis of perceived costs and benefits. Participants’ view of SPIRIT’s value proposition related to: 1. Utility - the content promised to be relevant and applicable, addressing current or future needs. Knowing SPIRIT was locally tailored increased expectations of utility. 2. Stimulation - content promised to be interesting. Presenters with “big names”, expert roles, and very senior policy experience piqued interest. 3. Persuasive marketing – clear communication using agency-attuned language that emphasised the value of SPIRIT, framed it in relation to agency values and goals, and was disseminated through locally appropriate channels. 4. Forecasting – the perceived quality of each intervention activity was used as an indicator of the likely quality of further activities, but only where participants were aware that they were all part of SPIRIT. The value proposition differs from M1 in that it was assessed in relation to each activity - the premise of SPIRIT might be rejected but individual workshops could still promise value.</td>
<td>Where managers saw the value proposition they espoused SPIRIT and encouraged participation, appearing genuine in their efforts. Liaison people supported SPIRIT based on (a) the extent to which it seemed likely to benefit their agency, and (b) whether acting as the LP would benefit or disadvantage them personally. Where liaison people saw the value proposition they went the extra mile to ensure the agency benefitted. Where staff saw the value proposition, participation was moved higher in their list of priorities. The potential benefits of SPIRIT counteracted the burden of data collection (“survey fatigue”) for agencies that had several data collection points prior to the intervention. Initially, SPIRIT marketing was suboptimal: dense, confusing, with poorly attuned “researchy” language (e.g. jargon and acronyms) – policymakers couldn’t see the value proposition. Strategic advice and input from liaison people improved communications substantially.</td>
<td></td>
</tr>
</tbody>
</table>

Most interviewees in A1, A5 and A6 saw potential value in SPIRIT and so were receptive and inclined to participate. In A3, persuasive internal marketing increased the value proposition from a lower base. Some managers in all agencies encouraged their staff to participate based on potential value. A2 and A4 generally saw less potential value in SPIRIT, and their liaison people put less effort into shaping and promoting it. Some staff in all sites were confused about SPIRIT’s purpose and form, did not know what was expected of them, and entangled the intervention with the trial. A particularly poorly received introductory session in A2 appeared to have lasting effects on perceptions of SPIRIT as a whole, despite some very well received workshops that followed.

---

delivered – but activities were not always as interactive or as participant-driven as intended. Authentic in-person leadership support and committed liaison people were vital mediators, while obstacles included confusion about the purpose of participation in SPIRIT, perceptions of poor alignment with agency practices or priorities, and feeling misunderstood or judged. Previous organisational change initiatives and archetypal views of researcher-policymaker relations sometimes appeared to underpin expectations and frame some of the concerns. The data collection demanded by the stepped wedge evaluation was onerous, and aspects of the trial were often entangled with participants’ perceptions of the intervention. Like many others, we found that pre-existing positive relationships between the agency and those involved in designing and implementing the intervention had considerable facilitative effects [75–77]. In our case, they helped to activate mechanisms such as respect and confidence.

**Implications for intervention improvement**

Given their pivotal importance, greater upfront engagement with each agency’s leadership and the nominated liaison person would have been beneficial. Local tailoring and shared decision-making was essential, but challenging for both the agency and the intervention team. For example, it

---

**Table 5** Mechanism 4 – “Getting good stuff”

<table>
<thead>
<tr>
<th>Context</th>
<th>How did mechanism 4 function?</th>
<th>How did mechanism 4 generate process effects?</th>
<th>Agency comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policy work tackles complex problems in complex systems. Views about the value of research are contested. Some policy staff were not interested in new skills or ideas, but most wanted and actively sought them. Many were unaware of existing resources in their agencies (human and technical). Each agency had distinct learning norms. Time pressures and reactive practices limited opportunities for policymakers to engage with new ideas and reflect on their practice. SPIRIT demanded considerable amounts of data collection (mostly because of the trial).</td>
<td>Participants felt they got good stuff from SPIRIT in relation to: 1. Utility - content was relevant, applicable, pitched at the right level, partly because presenters had “done their homework”. 2. Credibility - presenters with content and practice expertise “got it”, i.e. they understood the constraints and opportunities of policymaking and the need for pragmatism. 3. Tangibility - targeted case examples and problem-solving activities made concepts concrete. 4. Stimulation - dynamic presenters captured interest and imagination e.g. via compelling behind-the-scenes anecdotes. 5. Linkage - interactive activities connected participants to external experts and existing internal resources, and forged intra-agency connections by alerting them to colleagues with expertise or shared work agendas. 6. Learning congruence activities leveraged preferred learning styles. 7. Reflective space - workshops provided opportunities for critical thinking. 8. Orientation - new staff found that workshop discussions provided insights into how their colleagues view, access and use research.</td>
<td>The perceived ‘return on investment’ of participation encouraged (or discouraged) continued engagement in the intervention and data collection, positive word-of-mouth, and receptivity to SPIRIT’s ideas and resources. Feedback form data indicated very positive views, but several interviewees found some content irrelevant, applicable or boring and were less inclined to participate in other activities, including data collection. Some spoke negatively to colleagues about their experience, possibly influencing their decision to participate.</td>
<td>Very high numbers of participants across all agencies got good stuff from the majority of workshops. Feedback form results in all sites were extremely positive. Most interviewees engaged with some content, including those who did not anticipate value, and many identified “specific realisable benefits”. Some interviewees believed they would put ideas into practice, and a few had done so, especially newer staff for whom SPIRIT was “formative”. Interview data suggests fewer participants in A2, A3 and A4 got good stuff, but there are only minor differences in their feedback form data compared to other agencies (Additional file 2).</td>
</tr>
</tbody>
</table>
was often difficult for agencies to make strategic use of processes that they had not initiated such as trialling the services for commissioning research. Advice from agencies about how tailoring could be best supported in their context may have been beneficial, but the process of tailoring will always demand time and effort. This reflects the underpinning need for agency leaders to be committed to participation from the start.

Despite being selected for broad similarities, the six participating agencies had markedly different remits, practices and conceptualisations of evidence. SPIRIT’s audit and feedback process was effective in developing a shared understanding of each agencies’ current and desired research use capabilities, but better understanding of their practice goals and values, and greater collaboration in designing the intervention and data collection instruments (which every agency desired) could have sharpened the meeting of minds about what was needed and how to address it. Understanding what participants think about intervention goals, and using their ideas about what should be done in order to achieve those goals, is usually critical for success [78].

As noted previously, the realist distinction between intervention activities and mechanisms is crucial for theory-driven evaluation, but it is equally crucial in the development of context-sensitive intervention design and

---

**Table 6** Mechanism 5 – Self-efficacy

<table>
<thead>
<tr>
<th>Context</th>
<th>How did mechanism 5 function?</th>
<th>How did mechanism 5 generate process effects?</th>
<th>Agency comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>As analytic thinkers, many (but not all) participants were keen to critique their own, their programs’ and their organisations’ research engagement. Levels of confidence varied: audit survey data suggested the majority of staff lacked experience and confidence in accessing, appraising and using research, whereas 2/3 of interviewees said their skills were reasonable (many had research qualifications and/or experience) but, in most cases, could be improved. Some were apprehensive at the excessive standards of research use that SPIRIT might require. The quality of leadership feedback about staff and agency performance varied.</td>
<td>Self-efficacy refers to people’s beliefs about their capability to perform tasks and achieve goals * Their self-efficacy was activated by: 1. Pragmatism - presenters advocated realistic “good enough” goals which assured participants they could achieve acceptable practice standards in using research. 2. Affirmation - participants felt they were building on well-established capabilities partly due to strengths-based audit and mid-way feedback, plus sensitive facilitation in workshop activities, and leaders conveying confidence in their staff. 3. Modelling - high profile experts recounted ‘war stories’, successes and hard-won lessons that countered idealism and echoed messy local attempts to solve problems using research. 4. Experiential learning - trying out tools and systems increased understanding and confidence. 5. Demonstrating expertise - interactive activities enabled participants to contribute valuable local knowledge and skills to SPIRIT content.</td>
<td>More interviewees in A1, A5 and A6 than in the other agencies said they were encouraged by SPIRIT. In all agencies, the audit feedback largely supported managers’ understanding of, and confidence in, their staff’s capacity. A minority of participants with less confidence felt overwhelmed (e.g. a very highly rated evaluation workshop in A5 caused a few participants to see evaluation as outside their ability), whereas some others already had high self-efficacy (in all agencies, but possibly A3 in particular). Mid-intervention feedback that showed progress, and which may have supported organisational-level self-efficacy, was not always disseminated within the agency. A few interviewees in A1, A3 and A5 felt that agency-level participation in SPIRIT indicated that their CEO lacked faith in staff capabilities.</td>
<td></td>
</tr>
</tbody>
</table>

---

  
implementation planning. An intervention cannot simply ‘do’ respect, or ‘deliver’ self-efficacy, it cannot control the perceived attractiveness of its premise, or make internal facilitators act strategically. Activating these mechanisms is an evolving work-in-progress shaped by personalities, relationships and complex shifting environmental opportunities and constraints. Greater understanding of the mechanisms that generate desired (and undesired) process effects provides helpful guidance, but putting this learning into practice takes creativity, humility and reflexivity.

**Our contribution**

These findings add to the existing knowledge by surfacing evidence about how policymakers perceived and engaged with different aspects of an intervention trial designed to increase the extent to which they use research in their work. Our realist process evaluation approach goes beyond questions of implementation fidelity and ‘what works?’ to provide a more nuanced and theoretically informed account of how the intervention produced process effects, and why there was such variation across the six policy agencies.

As per Fig. 2, we anticipate that the intervention’s process effects, and the mechanisms that underpin them, mediate the study outcomes, but we caution against assumptions that this is a linear predictive relationship. As realist evaluation adherents indicate, there are usually multiple causal pathways in real world interventions, and the best we can do is identify common pathways for particular groups of individuals in particular circumstances; therefore, we concur with McMullen et al. that, “there is not, nor can there ever be, a universal implementation model for complex interventions. Site-specific characteristics and realities need to be considered” [79]. However, this consideration need not start from scratch with each new intervention – we can develop an increasingly sophisticated understanding of the conditions that make these outcomes more likely in a given setting.

As Pawson argues, “evaluation science assumes that there will be some pattern to success and failure across...
interventions, and that we can build a model to explain it” [1]. We hope to have made a start in identifying these patterns in a form that will enable others to extrapolate and apply lessons to other interventions and contexts [1].

**Strengths and limitations of this process evaluation**

Using a realist approach enabled us to identify and test hypothesised causal mechanisms, evaluate the extent to which SPIRIT activated them, use this analysis to refine the programme theory, and identify areas of strength and potential improvement in the intervention and trial design. The identification of underlying causal mechanisms and the development of propositions enhances the utility and transferability of the findings [3, 80] and strengthens the general knowledge base by building on existing theories.

The thematic overview of the process evaluation data in Additional file 1, and the inclusion of informing theory in Additional file 2, provide ‘analytical trails’ that support the findings.

Triangulating different types of data obliged us to consider diverse points of view and increased the trustworthiness of our findings. As Wells et al. [9] note, “... evaluations need to incorporate multiple methods, multiple sources and multiple perspectives if they are to reflect the context of practice adequately”. We achieved this thanks to (1) the unusually generous appointment of a dedicated process evaluation researcher throughout the study, and (2) the length of the intervention (12 months) and its staggered delivery, which gave us considerable time in each agency to test hypotheses at different points in the intervention across six sites. However, we acknowledge this was an exploratory first step and the ideas are yet to be tested by others and in different settings; therefore, at this stage, our findings are only a rough indication of

**Table 8 Mechanism 7 – Confidence**

<table>
<thead>
<tr>
<th>Context</th>
<th>How did mechanism 7 function?</th>
<th>How did mechanism 7 generate process effects?</th>
<th>Agency comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>The trial required a lot of data collection: six measurement points using three measures, and a process evaluation. Local goal-setting was informed by audit findings from two of the measures. All agencies had previously endured disruptive change initiatives which, in some cases, had little perceived benefit. Many participants had considerable expertise in intervention research and evaluation. Some felt researchers were naïve about policymaking and there was a degree of scepticism about what a researcher-initiated intervention could offer.</td>
<td>Participants had confidence in SPIRIT when they regarded the intervention, and the trial, as: 1. <em>Valid</em> - (a) the audit and feedback data seemed robust thus (b) the goal-setting was well founded and (c) subsequent data collection promised to track meaningful change and provide useful findings. 2. <em>Trustworthy</em> - the SPIRIT team were seen to act transparently and non-judgementally, in good faith. Where there were positive pre-existing relationships between agency and SPIRIT staff it facilitated trust in the initial stages. 3. <em>Transparence</em> - it was clear what demands would be made on participants and how data would be used, including when and how outcomes would be communicated. 4. <em>Safe</em> - it was clear that reporting the trial would not compromise individuals or agencies. 5. <em>Effective</em> – the intervention strategies could generate meaningful change in their setting.</td>
<td>Confidence affected the extent to which targeted policymakers wanted to take part in intervention activities and data collection. Liaison people reported that the audit feedback increased leaders’ confidence in, and enthusiasm for, SPIRIT which they communicated to their staff. Participants who lacked confidence in the measures or intervention design, and the small minority who questioned the study’s integrity or safety, expressed discomfort about SPIRIT. Some were avoidant and spoke poorly of researchers. Negative views of the trial were often entangled with the intervention. It seemed that scepticism about researcher-developed content seldom prevented participation, and follow up interview data suggested that attendance at a workshop tended to increase confidence.</td>
<td>45 leaders took part in audit feedback, and 38 completed evaluation forms. 37 of these (the exception was in A3) answered ‘Yes’ to the following statements: 1. The forum provided clear and accessible information. 2. It provided useful feedback on how we currently use research. 3. The presenter had appropriate knowledge and skills. 4. It gave me confidence that SPIRIT will be tailored to suit this agency. 5. I will encourage my staff to participate in SPIRIT. Despite this high level of confidence, information about the audit was often not disseminated effectively throughout agencies, thus many interviewees did not understand how it had been interpreted or how subsequent data would be used. In A1, a small minority of interviewees initially felt SPIRIT might threaten them or their agency with exposure, but some in all agencies expressed discomfort in “being researched”. Some in A2 questioned the integrity of the study, suggesting it was a business endeavour disguised as research. Liaison people and leaders were instrumental in addressing concerns and increasing confidence, but in A4 and A6 the liaison people themselves questioned the sensitivity of the measurement instruments.</td>
</tr>
</tbody>
</table>
major causal patterns within SPIRIT’s engagement and participation. Further testing and refinement are required.

A limitation was our inability to determine the full range of views and experiences of targeted staff in each agency. Interviewees were sampled purposively for maximum variation of relevant views and experiences, but many declined interviews and it was not always possible to identify substitutes. Others have found similar problems [52]. Consequently, we reached a smaller range of participants than envisaged and so may have missed important views. For example, all the process evaluation interviewees in A4 (11 people with a total of 15 interviews over the duration of the intervention) were either lukewarm or dismissive of SPIRIT, but during outcome measures interviews some A4 participants stated that they welcomed the intervention, and following the trial their CEO said SPIRIT had impacted his agency positively. In all agencies, we saw some non-agreement between the highly positive feedback form data and the more critical responses in the interview data. This may be the result of different foci – interviews ranged across the whole of SPIRIT (including its premise, communication and data collection), while feedback forms were workshop-specific – but other factors could be skewed sampling, leading interview questions or the bluntness of the feedback form. The response rate for feedback forms was good, with 74% of attendees completing them, but it is unclear whether those who did not complete forms differed from those who did, and thus what views we might have missed. The direction of this quantitative data was consistent with patterns in the qualitative data regarding a more positive response from agencies 1, 5 and 6, but feedback form responses across agencies and items were so similar that it is likely that the tool discriminated poorly. We used Yes/No statements to maximise response rates from participants who might be rushing to leave, but this was probably too limiting. Certainly, there were many occasions where the free text fields conveyed ambivalence or, at least, scope for improvement, when the scored statements suggested 100% satisfaction. We would use a more sensitive instrument in the future.

Table 9 Mechanism 8 – Persuasive leadership

<table>
<thead>
<tr>
<th>Context</th>
<th>How did mechanism 8 function?</th>
<th>How did mechanism 8 generate process effects?</th>
<th>Agency comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participating agencies were bureaucracies with strict hierarchies, but very different infrastructures and numbers of staff. Policymakers in the six agencies had varying levels of respect for their leaders, including different views about the extent to which their leaders valued using research. The professional behaviours of agency leaders in relation to SPIRIT were interpreted locally.</td>
<td>Leaders (managers, opinion leaders and liaison people) were persuasive in support of SPIRIT when: 1. they engaged in value messaging, i.e. they articulated the benefits of SPIRIT including their agency’s scope for and need for increased capacity in using research (M1) in a manner that did not demean current capacity (M5). 2. Respected managers modelled engagement with the interventions’ ideas and resources, thereby connecting SPIRIT to organisational values and priorities. When impressive expert presenters modelled their commitment to research-informed policy and provided examples of its benefits they positioned SPIRIT’s values in the wider policy environment. 3. Agency leadership support was visible; credible (these managers were known to support and engage in research-informed policymaking themselves); consistent (leaders across the agency conveyed support); and authentic (support seemed genuine). In-person advocacy and participation in workshops was experienced as more authentic than email espousal - actions speak louder than words.</td>
<td>Persuasive leadership probably impacted all the other mechanisms and process effects, and was a process effect in its own right (see section on mechanism interactions and feedback below). For example, it connected SPIRIT to organisational priorities (M1), increased perceptions of potential value (M3) and trustworthiness (M7), gave staff a positive message about the relationship between SPIRIT and current capabilities (M5), and provided liaison people with a mandate for action (M9). All these impacts, including expert presenters’ advocacy for pragmatic research-informed policy, increased receptivity to the intervention.</td>
<td>Leadership support was strongest in A1 and A5. In other agencies managers either lacked credibility as research advocates or their espousal appeared to be perfunctory or inconsistent (e.g. the CEO supported SPIRIT but some managers voiced dissent). Some presenters had insufficient policy-savvy credibility to function as leaders – in all agencies to a small extent, but it was ‘felt’ more in A2 and A4 where resistance was higher. In both A3 and A5 managers’ attendance in workshops bolstered perceptions of SPIRIT’s importance but also inhibited frank discussion by more junior staff. This may have occurred to some extent in other agencies. Forums targeting agency leaders were especially well attended when scheduled within or in lieu of formal management meetings.</td>
</tr>
</tbody>
</table>


Page 16 of 20
Table 10 Mechanism 9 – Strategic insider facilitation

<table>
<thead>
<tr>
<th>Context</th>
<th>How did mechanism 9 function?</th>
<th>How did mechanism 9 generate process effects?</th>
<th>Agency comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPIRIT required that each CEO nominate a local ‘liaison person’ to coordinate SPIRIT in their agency. Liaison people and managers had their own views about using research in policy work, and about the value of SPIRIT’s goals and strategies. In most cases they did not have a say in the agency-level decision to participate in SPIRIT. Beliefs about divisions between the ‘worlds’ of research and policy had currency.</td>
<td>In order to be strategic and facilitative, the internal coordination of SPIRIT depended on: 1. Translation - liaison people and managers used their ‘insider’ expertise to explain SPIRIT in terms that made most sense to targeted policymakers, illustrating how the intervention intersected with and complemented organisational goals and activities. 2. Mediation - liaison people actively identified concerns and worked with the SPIRIT team to resolve them. 3. Persuasive marketing - liaison people used local communication channels and creative strategies to “sell” SPIRIT (see M3). 4. Negotiation - liaison people and, in one agency, managers, advocated forcefully for adaptations to the intervention that would better suit their needs. 5. Support - liaison people were supported internally by managers and externally by a responsive SPIRIT team who provided materials, information and feedback (M6).</td>
<td>Like M8 above, strategic insider facilitation strongly affected other mechanisms as well as SPIRIT’s process effects. High quality facilitation fostered consultative tailoring, continuous engagement and informed participation. Information reached targeted staff in a form that was persuasive and accessible. Problems were identified and resolved so there was greater confidence and receptivity. Poor facilitation led to suboptimal tailoring, confusion about the intervention’s purpose and form, and a poorer local value proposition (M3), resulting in lower levels of participation or unwilling attendance.</td>
<td>Liaison people in A1, A3, A5 and A6 used creative strategies to tailor and champion SPIRIT, disseminating information through formal and informal channels, e.g. by nominating colleagues to give updates at their team meetings and promoting forthcoming events over coffee. There were transitional lags during changes of liaison person in A3, A4 and A6. The appointment of an enthusiastic and well-supported liaison person in A3 increased survey response rates. Most liaison people spoke highly of the SPIRIT team’s support, but A6 experienced too many points of contact and A4 said there was insufficient guidance. As noted, the liaison people in A2 and A4 did not facilitate SPIRIT to the best of their abilities because they did not believe it was worthwhile (M1, M3, M7). This probably contributed to a lower value proposition (M3) for targeted participants in those agencies, and less confidence (M7). The perceptions, behaviour and impact of liaison people is covered in more detail elsewhere.</td>
</tr>
</tbody>
</table>


Table 11 Initial and revised programme theory

<table>
<thead>
<tr>
<th>Initial programme theory (a-contextual)</th>
<th>Revised programme theory (contextually contingent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPIRIT will engage and motivate agency leaders to ‘own’ the intervention using audit feedback, deliberative goal-setting and programme tailoring – this agency-driven approach will generate a priority-focused programme that offers locally relevant practice support and accommodates differences in agencies’ values, goals, resources and remits. The programme will comprise a suite of andragogical activities, tools and connections across the research-policy divide that provide resources and build knowledge, skills and relationships; and will be supported via modelling and opinion leadership by agency leaders and dynamic external experts. CEOs will promote SPIRIT in their agencies and liaison people will facilitate the tailoring and implementation – these strategies will act synergistically to stimulate and resource participants at different organisational levels, leading to changes in values, practice behaviours and agency processes. This will facilitate increased use of research in policy processes.</td>
<td>Where agencies have an existing orientation to use academic research and are on a trajectory of improved use with perceived room for improvement, SPIRIT will be used to complement or trigger organisational initiatives. Where liaison people and agency leaders believe in the value of the intervention and have confidence in the measures, they will play a pivotal role in tailoring the intervention and championing its goals. Leaders will be motivated by deliberative audit feedback and goal-setting. In all sites, ownership will be increased by greater consultation, collaboration and choice. Agency-attuned communications will be vital in explaining goals, conveying value and addressing concerns. Andragogical activities, tools and connection across the research-policy divide will be valued in all agencies where they leverage existing strengths and address local concerns pragmatically. Staff will make use of these opportunities where they see concrete benefits, and newer staff may benefit most.</td>
</tr>
</tbody>
</table>
Reflections on conducting a realist process evaluation

Conducting a realist process evaluation was immensely valuable, but time consuming and challenging. Like others (e.g. [49, 81]), we struggled to disentangle aspects of the causal pathways; specifically, to delineate mechanisms from intervention strategies, contexts and outcomes. Realist analysis does not have a step-by-step guide, and it presents a unique tension between ontology and epistemology, so we sometimes struggled to reconcile our search for factual existing mechanisms with the need to take an “imaginative leap” and postulate those mechanisms [82]. Three strategies helped: first, scanning appropriate literature and drawing on established theories, for example, the concept of relative advantage [6, 58, 83] was critical for understanding variation in perceptions of SPIRIT and how this linked to the communication strategy. Second, the realist emphasis on counterfactual thinking [54] was very helpful in weighing up the plausibility of different theories. Third, reminding ourselves that causality does not function as discrete components or configurations and that our analysis was intentionally abstracting for the purposes of theory building rather than attempting to depict reality in all its messy, interdependent glory (see also [61]).

Conclusion

This realist process evaluation describes how participants experienced different aspects of a multi-component research utilisation intervention in policy organisations, and why there was such variation across the six implementation sites. We identify nine mechanisms that appeared to facilitate engagement with and participation in the intervention in these settings: (1) Accepting the premise (agreeing with the study’s assumptions), (2) Self-determination (participative choice), (3) The value proposition (seeing potential gain), (4) ‘Getting good stuff’ (identifying useful ideas, resources or connections), (5) Self-efficacy (believing ‘we can do this!’), (6) Respect (feeling that SPIRIT understands and values one’s work), (7) Confidence (believing in the study’s integrity and validity), (8) Persuasive leadership (authentic and compelling managerial advocacy) and (9) Strategic insider facilitation (local translation and mediation). This analysis was used to develop tentative propositions and to revise the overarching programme theory. Although our findings are nascent and require further testing and refinement, they indicate areas of strength and weaknesses that can guide the development and implementation of similar studies in other settings, increasing their sensitivity to the range of issues that affect the value and compatibility of interventions in policy agencies.

Additional files

Additional file 1: Descriptive overview of results. (PDF 403 kb)
Additional file 2: Supporting theory. (PDF 335 kb)
Additional file 3: Summary of SPIRIT intervention implementation fidelity. (PDF 1460 kb)

Abbreviations

CEO: Chief executive officer; SPIRIT: Supporting Policy In health with Research: an Intervention Trial

Acknowledgements

Many thanks to the agencies and individuals who participated in SPIRIT. We are indebted to the policymakers who took time out of busy schedules to contribute to our process evaluation. Thanks also to Emma Darsana for help with data management and coding; to Huw Davies, Stacy Carter and Joanne McKenzie, who offered advice on previous versions of this paper; and to the reviewers who gave us constructive feedback.

Funding

SPIRIT was funded as part of the Centre for Informing Policy in Health with Evidence from Research (CIPHER), an Australian National Health and Medical Research Council (NHMRC) Centre for Research Excellence (#1001436), administered by the Sax Institute. CIPHER is a joint project of the Sax Institute; Australasian Cochrane Centre, Monash University; University of Newcastle; University of New South Wales; Research Unit for Research Utilisation, University of St Andrews and University of Edinburgh; Australian National University; and University of South Australia. The Sax Institute receives a grant from the NSW Ministry of Health. The Australasian Cochrane Centre is funded by the Australian Government through the NHMRC. AH is supported by an NHMRC Public Health and Health Services Postgraduate Research Scholarship (#1093096).

 Authors’ contributions

AH led the process evaluation design, data collection and analysis, and drafted the manuscript. SB contributed to the design, independently conducted some analyses and, together with AH, workshopped initial findings to refine them. SM led the statistical analyses. All authors made substantial contributions to the analysis and interpretation of data, and were involved in critically revising the manuscript for important intellectual content. All authors read and approved the final manuscript.

Authors information

Not applicable.

Ethics approval and consent to participate

All participants provided informed consent. Both the agencies and individual staff members were advised they were free to decline to participate in any or all aspects of SPIRIT at any time without explanation. Ethical approval was granted by the University of Western Sydney Human Research Ethics Committee, approval numbers H8855 and H9870.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

1Sax Institute, 235 Jones Street, Ultimo, NSW 2007, Australia. 2Sydney School of Public Health, Edward Ford Building (A27), University of Sydney, Camperdown, NSW 2006, Australia. 3Australasian Cochrane Centre, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, Victoria 3004, Australia. 4School of Medicine, University of Notre Dame, 160 Oxford St, Darlington, NSW 2010, Australia. 5Centre for Medical
References


Appendix 14. Descriptive overview of results

Mechanism 1. Accepting the Premise

To be perceived as relevant and potentially beneficial, SPIRIT had to establish its compatibility with each agency’s remit, practices and current trajectory of change. In half the agencies the intervention was regarded as a good fit that completed existing or planned initiatives—“it reinforced what we were trying to do anyway”—and many interviewees saw SPIRIT as addressing “a real need”. This translated into some enthusiasm for the intervention, particularly among groups of participants in A1 and A5 who were looking for ways to improve the use of research in their practice. In two agencies the liaison people and many of the other participants rejected the premise of SPIRIT because there was a perceived disconnect between the intervention’s assumptions and the agency’s values; specifically, differences in their conceptualisations of evidence and how it should be developed (A4), and differences in their beliefs about how interventions should be designed and implemented, which amounted to feeling that SPIRIT was imposed and too prescriptive (A2). The agencies’ values were bound to the needs and preferences of their primary stakeholders, which for A2 and A4 meant that clinicians’ views and ‘bottom-up’ initiatives had primacy. A3 struggled to accept the premise of SPIRIT for different reasons. They self-identified as having such high standards of research use that, according to one manager, there was “no room for improvement”. This view was echoed by other staff who were surprised their CEO felt participation in a research utilisation intervention was warranted.

Even where they saw substantial scope for improvement, some staff in all sites questioned the need to address their use of research as a priority, “I don’t think there’s a need for any kind of urgent remedial action here”. In A2, some interviewees argued that they should not have to improve their ability to access and appraise research, believing that research experts should be doing this work, “why can’t they just do it for me rather than tell me how to do it?” One interviewee admitted she was not using research in her work (despite plenty of scope and the availability of relevant studies), yet rejected the need for SPIRIT because “We’re doing it already”. There was a sense that their use of research was good enough for their purposes, and that they would prioritise other issues if they wished to improve practice.

Surprisingly, Accepting the Premise was not a universal deal breaker. It had profound impacts on liaison people (see below), but many other participants who rejected the premise of SPIRIT appeared to assess each intervention activity on its own merits—possibly because they were unaware that they were part of SPIRIT—and often reported that they found workshops worthwhile. We do not know if, or how many, potential participants failed to take part because they had dismissed the intervention as incapable of benefitting them or their agency.

This mechanism seemed to have a more fundamental impact on liaison people who were required not only to administer the study but to attach themselves to it as champions. In the two agencies where they rejected the premise it profoundly influenced how they tailored, administered and promoted SPIRIT (M9). Rejecting the premise also appeared to reinforce beliefs about a generalised disconnect between researchers and policymakers, and to fuel scepticism about the agenda of research-informed policymaking, “I just think it probably just reinforced to me that what we do is a lot more complex than perhaps how academics can describe it.” Acceptance of SPIRIT’s premise was bolstered where there was advocacy support from management (M8) and strategic internal facilitation (M9). The deliberative conversations held with managers as part of the audit feedback galvanised leadership support and helped to calibrate the intervention with each agency’s needs, but was not always sufficient to address fundamental concerns about the underlying premise.
Mechanism 2. Self-determination

This mechanism incorporates interactions that enabled participants to regard SPIRIT as empowering and self-driven rather than constraining and imposed. Self-determination was a process as well as a destination, and is strongly linked to M6: Respect.

As intended, the intervention’s value was undoubtedly enhanced by its flexibility, enabling agencies to identify local goals and tailor content for increased relevance and applicability. Policymakers naturally want to be treated as experts in their domain, and the audit feedback forums were highly successful in recognising and building on this. This contributed to a sense of ownership in four agencies but, despite positivity about the audit feedback deliberations (which were very well received in all agencies), some leaders in A2 and A4 argued there was insufficient flexibility to use SPIRIT in a way that suited them, and that fixed options contributed to poor compatibility. Given that the form of SPIRIT was not flexible enough for its premise to be customised, tailoring was only embraced in agencies where the exiting premise was acceptable (M1). Tailoring was also onerous: agencies often struggled to match SPIRIT options to their needs, and decision-making was complicated by bureaucratic processes so that some liaison people had to develop new pathways for making and authorising decisions about customisation. Liaison people and managers put very different levels of effort into this consultation and customisation, but where managers and, particularly, liaison people actively engaged in tailoring, they shaped SPIRIT more than anticipated. For example, A6 insisted on greater customisation than was initially allowed, A5 negotiated ‘extras’, and A3 modified the participant eligibility in their setting. Where liaison people consulted widely, interviewees generally saw the intervention as more attuned to their needs.

Opportunities for self-determination affected how people engaged in and perceived workshops. The expert presenters were asked to make workshops highly participative, but interactivity varied hugely. At its best, participant contributions shaped the content via questions, case examples and robust debate. At worst, participants felt “lectured at” (which also negatively triggered M6: Respect). Although they were encouraged to co-present workshops, only one agency took this up. This workshop was very well received, but since most of the workshops were very well received and the feedback form instrument discriminated poorly, it is not possible to determine how much difference co-presentation made.

A few interviewees, particularly in A1 (which was most involved in policymaking rather than program development) worried that SPIRIT was critical of ‘craft’ practices and might assert a protocol dictating how research should be used that would “infringe on the art of writing policy”. They wanted the freedom to use research flexibly, as they saw fit. Some participants felt obliged to attend workshops and complete measures due to managerial expectations or, in one agency, explicit demands. A few liaison people took on the role reluctantly, but this probably only affected implementation in A2 and A4 as the other ‘put upon’ liaison people were not involved in the active phase of SPIRIT.

Trial demands surrounding the intervention affected self-determination. The intervention’s randomised start date was inconvenient for some agencies (so much so that one agency insisted on a postponement). In several sites, participants felt their agency could have used the intervention more strategically if they had been able to determine when it took place. Most interviewees would have liked the whole-of-study to have been collaborative, arguing that the intervention form and content, and the data collection instruments, would have been better attuned to their agency as a consequence.

Mechanism 3. The Value Proposition

A “value proposition” (promised advantage) is a convincing argument about the worth of a strategy that is assessed by prospective users on the basis of perceived costs and benefits. [1] In the case of SPIRIT, the intervention as a whole and its individual activities were judged on their value proposition. This judgement affected receptivity to the intervention’s ideas. Most participants saw potential value in SPIRIT but were “agnostic” prior to the intervention’s start. They invested time based on anticipated utility—“concrete payoff”— and, to a lesser extent, interest. They were attracted by high profile presenters with “big names”
and impressive biographies; persuasive promotion of SPIRIT activities by managers and the liaison person; and knowing that content had been tailored for them. The mix of presenters from research and policy fields was valued by most, but some had low expectations of researchers, “I just think, oh, academics, we’re going to get talked at and get a lot of overcomplicated things, not in plain English”. Value was often deferred, many individuals commented that they had little to gain from SPIRIT themselves, but that others (more junior or less experienced colleagues, or the agency’s executive) could benefit.

Communication was crucial in establishing the value proposition. Participants in all agencies wanted five questions answered: 1. What’s in it for us? 2. What do we have to do and when? 3. Why are we doing it? 4. Who are you? and 5. What are you doing? However, it took some time before the SPIRIT team caught up with this. Early communication was poor, causing confusion and some alienation. SPIRIT’s acronyms were universally loathed. Some participants’ beliefs that researchers and policymakers operate on different planes were exacerbated by dense, jargon-laden information. Consequently, many participants were unsure why their agency was participating in SPIRIT, “there is a bit of confusion... ‘What are we doing this for?’”. In A1, managers used a key phrase from a workshop—“using the best available research in the time available”—to encapsulate their goals in participating and their philosophical stance in relation to research more generally. But other agencies struggled to identify exactly what they hoped to get out of SPIRIT. Ultimately, persuasive marketing was dependent on acceptance of SPIRIT’s premise (M1).

Communication was also critical in minimising the threat that trial demands posed to the value proposition. For example, poor understanding of SPIRIT undermined confidence (M7): the six measurement points had been explained but remained puzzling for many, partly because of the non-intuitive rationale for multiple measurement points (unlike before and after measures). This led to “survey fatigue” and a perception of redundancy: “I wonder why we had to have two baselines?”. Some participants thought data collection interviews were part of the intervention. Where data collection was framed by liaison people as part of the value proposition (in that it by provided performance indicators or operated as part of SPIRIT’s quid pro quo), response rates increased. Where agencies had multiple data collection points prior to the intervention (because they had been randomised to a later start date) holding strategies were required to emphasise forthcoming returns on investment.

Mechanism 4. ”Getting Good Stuff”

Feedback form data indicated that workshops were extremely well received (Table 1). Across all agencies, almost all participants judged that the workshops were interesting, and had presenters with appropriate knowledge and skills, and reported that SPIRIT was likely to be beneficial to their agency (items 1, 4 and 6: ≥ 96% of responses across all agencies). A slightly lower, but still high, proportion of responses indicated that participants expected to use information from the workshops in their work (item 5: ≥ 93% of responses across all agencies). Perceptions of the workshop’s relevance (item 2) and the extent to which it was realistic about the challenges and constraints of the agencies work (item 3) were the only items where any potentially important differences between agencies were observed. Perceptions of relevance were marginally lower in A2 compared to the other agencies (followed by A4), and the extent to which the workshops were regarded as realistic about the agency’s work were marginally lower in A4, followed by A3.
Table 1. ‘Yes’ responses on SPIRIT workshop feedback forms: total numbers and percentages

<table>
<thead>
<tr>
<th>Feedback form statement</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>All agencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The workshop was interesting</td>
<td>77</td>
<td>92</td>
<td>82</td>
<td>66</td>
<td>103</td>
<td>71</td>
<td>491</td>
</tr>
<tr>
<td></td>
<td>(100%)</td>
<td>(97%)</td>
<td>(98%)</td>
<td>(96%)</td>
<td>(98%)</td>
<td>(100%)</td>
<td>(98%)</td>
</tr>
<tr>
<td>2. The workshop was relevant to my work</td>
<td>73</td>
<td>86</td>
<td>81</td>
<td>63</td>
<td>97</td>
<td>72</td>
<td>472</td>
</tr>
<tr>
<td></td>
<td>(95%)</td>
<td>(90%)</td>
<td>(96%)</td>
<td>(91%)</td>
<td>(92%)</td>
<td>(100%)</td>
<td>(94%)</td>
</tr>
<tr>
<td>3. The workshop was realistic about the challenges and constraints of our work</td>
<td>41</td>
<td>42</td>
<td>23</td>
<td>35</td>
<td>65</td>
<td>56</td>
<td>262</td>
</tr>
<tr>
<td></td>
<td>(100%)</td>
<td>(93%)</td>
<td>(89%)</td>
<td>(83%)</td>
<td>(94%)</td>
<td>(98%)</td>
<td>(94%)</td>
</tr>
<tr>
<td>4. The presenter had appropriate knowledge and skills</td>
<td>83</td>
<td>101</td>
<td>90</td>
<td>75</td>
<td>110</td>
<td>76</td>
<td>535</td>
</tr>
<tr>
<td></td>
<td>(100%)</td>
<td>(100%)</td>
<td>(98%)</td>
<td>(96%)</td>
<td>(99%)</td>
<td>(99%)</td>
<td>(99%)</td>
</tr>
<tr>
<td>5. It is likely that I will use information from this workshop in my work</td>
<td>41</td>
<td>32</td>
<td>67</td>
<td>53</td>
<td>81</td>
<td>51</td>
<td>325</td>
</tr>
<tr>
<td></td>
<td>(98%)</td>
<td>(94%)</td>
<td>(93%)</td>
<td>(96%)</td>
<td>(96%)</td>
<td>(94%)</td>
<td>(95%)</td>
</tr>
<tr>
<td>6. It is likely that SPIRIT will benefit my agency</td>
<td>33</td>
<td>24</td>
<td>57</td>
<td>46</td>
<td>74</td>
<td>46</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>(100%)</td>
<td>(100%)</td>
<td>(100%)</td>
<td>(96%)</td>
<td>(99%)</td>
<td>(96%)</td>
<td>(98%)</td>
</tr>
</tbody>
</table>

We observed small differences in the mean total feedback score across agencies, with the lowest overall scores reported in Agencies 2 and 4 (Figure 1). The maximum difference between any two agencies was 1.4 (out of 6), or the equivalent of a different response to one of six items. Agency 2 had lower score than all other agencies (ranging from 0.9 lower than agency 4 to 1.4 lower than agency 5). The difference in the mean total feedback score between other agencies was small (from 0.2 to 0.6 out of 6).

Figure 1. Total feedback form scores for each agency in relation to the six statements

* Significant differences are marked with a star

Many interviewees concurred with the positive feedback form responses, “They were some of the best interventions I’ve ever seen”. The use of passionate “service oriented” presenters with hands-on expertise who
had “done their homework” and who “got it”, together with compatibility with preferred learning styles, resulted in useful and “invigorating” content that participants hoped to apply. Targeted case examples and memorable insights provided by behind-the-scenes stories, especially those recounting hard-won lessons, were particularly valued. The opportunity to talk with researchers was appreciated when the topic was regarded as relevant and applicable, and four agencies planned further seminars or collaborations with a presenter. Although many interviewees felt their learning had been consolidated rather than advanced, this was usually seen as worthwhile. In agencies with unengaged liaison people (A2 and A4), interviewees viewed SPIRIT as having less value overall, but, despite rejecting the premise (M1), and in A2 feeling obliged to attend (M2), many participants reported that they found individual workshops useful.

Policymakers in all agencies valued the opportunity for reflection and to explore topics with colleagues across the agency—“it really brought people together”—which sometimes led to the identification of shared interests and internal resources. Newer members of staff used workshops as orientation, a chance to gauge alignment between agency norms and SPIRIT’s idea, while those without research expertise were most enthusiastic about SPIRIT providing valuable ideas and tools: “[SPIRIT happened] at the very start of me working here, so it’s been formative for me”. In A3, where staff saw little room for improvement, SPIRIT affirmed their view: “So we’re now walking around very conceited, with very big heads, thinking... we use evidence in the best way possible... [but] it’s nice to have it confirmed”. Several managers and liaison people identified value in forging or enhancing relationships with the SPIRIT team, a side effect of participation.

Where workshops failed to align with local learning practices (e.g. not distributing advance readings) they were regarded as having missed an opportunity and were considered less useful (A2 primarily). As anticipated, didacticism was unwelcome, but interactivity was hard to assert as some presenters did not follow their brief, and liaison people sometimes attempted to increase the value of workshops by maximising content at the expense of participation, which often appeared to have the opposite effect – participants felt overwhelmed and unable to take in key messages.

Perceptions of other intervention components were harder to access. Many interviewees were vague about whether they had seen CEO emails espousing SPIRIT (M8), or if they were receiving weekly updates about resources in the online portal. Those that had accessed the portal said they found it helpful (albeit cumbersome to access due to the need for a password), but could not identify any specific use. We did not manage to interview those involved in the brokered services in A3 and A4. In the other agencies, the response was mixed—only A5 and A6 were entirely happy about the final product and could identify ways that it would be used. Dissatisfaction mainly appeared to be an artefact of the trial: agencies found it difficult to identify which service and what topic would best meet their needs when required to do so within an externally imposed timeframe. Several participating agencies that struggled to select and tailor their brokered service had a history of using the service previously with high levels of satisfaction, but those occasions had been agency-initiated and thus needs-driven.

The six data collection points demanded by the trial tainted some participants’ view of the intervention’s value, but was usually judged to be “worth the hassle” due to the value of free high calibre workshops and other intervention resources.

**Mechanism 5. Self-efficacy**

Two-thirds of those who contributed to the process evaluation said they used research on a regular basis and had a reasonable degree of existing capacity, albeit with room for improvement in most cases. Some felt they had the capacity but that their current role or project did not require research input. A few felt they neglected opportunities to use research either because they lacked the skills and confidence, or because of time pressures and insufficient onus on research use in the organisational culture to warrant prioritising its use, “[we] don’t have time to sit around and ponder... [we have] to get on and do something”.

Some participants reported that SPIRIT workshops had raised the bar in terms of research appraisal or program evaluation, but in agencies where evidence was seen as more fluid and consensual (or driven by
“political imperatives”) there was little perceived need to improve their mastery of academic research utilisation. An example of this was the emphasis on stakeholder advice as core evidence in A2 and A4: “I think the culture is you just get up and go to the person that’s in our network and say ‘I’ve been asked to do something on this. Who shall I talk to or can you head me in the right direction?’.”

Nevertheless, participants identified several intervention strategies that they believed supported self-efficacy generally. Where workshops were explicitly pragmatic it gave permission for “good enough” practices that participants found motivating: “We can do this”. Many mentioned the message “Do the best you can with what you can get in the time available” (from the Leaders’ Forum run by Prof John Lavis) as especially helpful. As one of the liaison people put it, “the last Leaders’ Forum really galvanised people... it was practical, realistic... it was potentially technically challenging to get it in place but something that could be done.” Another participant talked about her confidence that she would be able to use resources provided by SPIRIT, “it gave the idea that once you’ve identified them implementing them actually isn’t that complicated. It’s sort of using the same skills that you’re already using... so you don’t need to have a PhD to be able to understand research and think about how you might apply it to your work.” Presenters’ forthright accounts of their struggles and successes in using research in policy acknowledged real world challenges and provided realistic goals. Participants valued opportunities to contribute expertise via discussion and, in one case, co-presenting. They particularly appreciated it when providers’ recognised that the agency was already research-engaged and had developed skills in integrating research with other forms of locally-relevant evidence. Workshops were also valued as a rare chance for critical reflection: “space to think about research and where it fits in”.

Affirmation by presenters, and via the audit feedback, showed agencies that they were building on well-established capabilities and were already part of the way there: “It was confirmation we’re heading in the right direction”. However, it was not possible to pitch workshops at the right level for all members of heterogeneous staff groups, so there were winners and losers. Some participants gained nothing new and, in some cases, those with little expertise felt overwhelmed. For instance, after an evaluation workshop an epidemiologist explained, “if you felt confident in [evaluation] that would be misplaced confidence... That’s what I took from that.”. Affirmation by leaders also bolstered confidence, but a few interviewees (mostly experienced, mid-level managers) inferred that agency participation in SPIRIT meant their CEO believed staff lacked competency in using research.

The SPIRIT team anticipated that the brokered services would provide useful experiential learning in commissioning research and evaluation, but most interviewees who were involved in that component of the intervention were already experienced in commissioning and did not find the process itself to be particularly valuable.

**Mechanism 6. Respect**

This was a fundamental mechanism that was anticipated by the designers and was successfully incorporated in most of SPIRIT’s activities. However, it had a profound effect when it was activated negatively, which the context made more likely: many participants evoked pre-existing views about “typical researchers” being both naïve and arrogant. Even if they did not subscribe to these views—in fact, many reported positive experiences of working with researchers—it was a pervasive cultural stereotype. In some cases it may have contributed to a conscious decrease in contact with researchers, for example, an executive in one agency explained that they had found university-based researchers to be “so far from reality” that the organisation tended to “stay away from academics more often than not”.

The extent to which SPIRIT was perceived as an attempt to infringe on the “art of policymaking” (discussed in Mechanism 2. Self-determination, above) was also seen as an indicator of respect. Experienced policymakers wanted recognition of their expertise.

Perceived lack of respect in aspects of SPIRIT caused resentment and, in some cases, withdrawal. Participants reported the following triggers: presenters making assumptions about or downplaying the
challenges of participants’ work; presenters failing to elicit participants’ input in workshops; failure by SPIRIT staff to be responsive or ensure participants were “kept in the loop” via regular feedback; feeling judged when being interviewed for the measures (in two agencies); patronising language and the perceived assumption that policy staff had poor understanding of research or needed to engage with research differently (some examples in all agencies); and the lack of options re shaping SPIRIT. In every agency, some participants felt SPIRIT had a deficit approach and called for a more inductive needs analysis that built explicitly on their practice strengths, but this was particularly pronounced in A2 and A4.

**Mechanism 7. Confidence**

Most participants, particularly in A1, A3, A5 and A6, expressed or implied confidence in SPIRIT, particularly where they had positive views of, or existing relationships with, the SPIRIT team, but some were sceptical about the sensitivity of the measurement instruments. This included two liaison people (who were responsible for facilitating data collection): one doubted the measures captured meaningful information about their practice, and the other felt they did not align with behaviours targeted by the intervention. For a few participants, these concerns undermined confidence in using audit findings to inform goal-setting. However, interviews with agency leaders and the liaison people clearly showed that the deliberative conduct of the feedback forum itself boosted confidence in the intervention. Evaluation forms supported this.

A number of interviewees expressed scepticism about SPIRIT’s ability to effect change. They argued that engagement with ideas and resources via two-hour workshops— no matter how inspiring the workshops were—was not enough to alter established practice. Two limitations were identified. First, SPIRIT’s limited ability to harness learning mechanisms with a predominantly workshop-focused intervention; the intervention was providing “refreshers” rather than training. Second, they pointed to organisational culture and current practice norms as powerful counter agents, and argued that change mechanisms had to be “embedded in our day-to-day processes” if practices were to change and be sustained. They recognised the role of organisational systems and leadership in making this happen.

A minority of interviewees across all sites expressed suspicion that data about individuals’ or teams’ performance might be used internally. This concern appeared to be sparked by experience of previous interventions and was more pronounced in agencies with less engaged liaison people, apparently because liaison people were actively addressing concerns in other agencies. Conducting the intervention as part of a trial added another layer of threat regarding potential public exposure; for example, the danger that publication of sensitive evaluation data might show the agency in a poor light, possibly by comparing it with other agencies participating in the trial.

A small minority of interviewees questioned the agenda behind the trial, “who will ultimately benefit?”, while some in A2 openly questioned the integrity of the study, “are you actually just gunning for business?”. Such concerns may have been exacerbated by the increasingly market-orientated culture that the agencies were operating in, and their own desire to maximise returns on investment from commissioned and partnered research rather than simply handing out funds.

Initially, some interviewees doubted researchers could develop policy-useful content, but participation in intervention workshops seemed to diminish this belief. Some participants expressed discomfort at “being researched” and a few blamed this on their agency’s CEO: “She [the CEO] determined she wanted to put us in a petri dish and compare us to others.”
Most concerns were satisfactorily addressed where liaison people worked with the SPIRIT team to actively identify and respond to them (M9), and authentic leadership support appeared to bolster participants’ confidence in SPIRIT overall (M8).

**Mechanism 8. Persuasive leadership**

Interviewees in all agencies reported that some managers encouraged participation in SPIRIT, but this was especially consistent in A1 and A5. Support was perceived as most genuine, and most memorable, when managers promoted SPIRIT informally and in-person. For example, interviewees usually did not recall the quarterly email endorsements from their CEO, or regarded them as corporate diplomacy, but they were impressed when senior managers made ad hoc expressions of support in conversations and meetings and, particularly, when they attended workshops: “seeing our directors engaged has meant that we’re seeing [SPIRIT] as more important”. Managers’ comments during activities also provided guidance about how to relate to the content, including what ideas were important and how they applied to agency practice. High profile external experts—known leaders in their field—modelled a broader commitment to research-informed policy.

Aware that their participation could increase the value of SPRIT, some managers in most agencies, including CEOs, attended workshops to model engagement (‘set a good example’). However, in A3 only one executive member participated, and only briefly. In the largest and most distributed agency, A6, workshop attendance was more strictly delineated so frontline staff were largely unaware that executive managers were participating in SPIRIT activities. Managers and the liaison person in one site undermined their CEO’s authentic endorsement of SPIRIT by showing disdain for the intervention’s goals and form. Thus visible leadership support for SPIRIT did increase perceptions of its value, but mostly where this support was consistent across the management team, in-person and demonstrably genuine. A downside was that, on occasion, managers’ presence in workshop inhibited frank debate, particularly where discussion focused on current weaknesses in research use.

**Mechanism 9. Strategic Insider Facilitation**

Liaison people and some leaders played a critical role in improving participants’ grasp of SPIRIT, using their insider knowledge to translate and maximise uptake of information. But the extent to which the liaison people accepted the premise of SPIRIT (M1) and were thus able to function as genuine champions varied considerably. Crudely speaking, in four agencies they championed SPIRIT and in two they did not. In agencies with less engaged liaison people, interviewees viewed SPIRIT as having less value overall, some staff did not know about SPIRIT activities and many were unaware that the intervention had been tailored for them. Conversely, where liaison people were most active they used creative strategies to promote upcoming activities and disseminate information through formal and informal channels, e.g. by nominating people to give updates at team meetings and encourage participation by their peers. And they functioned as mediators, ensuring that misconceptions and concerns were identified and addressed, and the SPIRIT team were advised of problems as they arose. Research team responsivity (see M6) was also vital for successful mediation. For example, liaison people could resolve complaints about the length of the online survey because the research team shortened it. The perceptions, behaviours and impacts of liaison people are explored in a dedicate paper published previously [2].

**Appendix 14. references**

Appendix 15. Supporting theory for the realist process evaluation analysis

Mechanism 1. Accepting the premise

To succeed, interventions (including research utilisation interventions [1]) have to address a real need, i.e. to offer greater benefits than current practice. [2-4] To do so they must be perceived as compatible with participants’ and organisation-wide’ beliefs, assumptions, values and work practices [3, 5] and tackle a pressing concern. [6] Staff will resist an intervention if it does not appear to be in the interests of their agency or its primary stakeholders [7], or if there is a belief that “we already do that”. [8] Organisational characteristics (e.g. absorptive capacity and receptivity to change [9]) affect how staff value evidence [1], and different conceptualisations of evidence will be meaningful in different contexts, e.g. ‘social proof’ (knowing through doing) may be more compelling than ‘scientific proof’. [10] Policymakers do not always trust the institution of research or find its outputs legitimate. [11] Audit and feedback theory posits that performance data which shows a gap between current and desired practice can establish the need, and create motivation for, change. [12] This requires credible audit data (M7) and shared goals.

Mechanism 2. Self-determination

Self-determination is the feeling of having control. [13] The value of involving participants in the design and implementation of interventions is well established – top down interventions seldom work. [3, 14-16] Eliciting participants’ views about the legitimacy of intervention goals, and the methods for achieving them, is critical for success goals. [17] New ideas and resources are more likely to be taken up if those they target “have sufficient opportunity, autonomy, and support to adapt and refine the innovation to improve its fitness for purpose”. [3] Adult learners want to self-direct their professional development and draw on experiential knowledge [18], so workshops must be flexible and participative. Where possible, participants should be involved in planning their learning. [4]

System-wide trends can encourage uptake of interventions [2] but if staff may feel obliged to take part [19] this can result in symbolic participation with minimal follow-through. There is an important difference between ‘towing the line’ and ‘deep engagement’. [7] Incentives can motivate participation, provided they are targeted appropriately [20]. Greater self-determination is positively associated with motivation to change [21] and ownership of change processes. [10, 22]

Mechanism 3. The Value Proposition

It is critical that participants are aware of and understand the intervention and its goals, and how intervention activities are likely to affect them [2, 3]; but multi-component interventions may be harder to understand. [23]. Policymakers want clear, succinct information, yet researchers often default to jargon. [24]. Participants will make judgements about the intrinsic value of the intervention based on how “the intervention is bundled, presented, and assembled”. [2] This “selling game” requires insider rhetoric and leadership support. [8]

As Karanika and Biron (2013) argue, “change, be it at the individual, organizational or societal levels, is about individuals making sense of and assigning meaning to events and their environment”. [25:241] This meaning, while made internally, is developed through comparison and negotiated with peers. [25, 26] It can be reframed, but it cannot be managed in a wholly predictable manner [3]—ideas travel through organisations in fluid ways [27]. The question of ‘What’s in it for me?’ and the ways that individuals express their views about value will affect peer enthusiasm or resistance [2] and overall engagement. [14]
Mechanism 4. “Getting good stuff”

Relevance and usability are recognised as critical features in intervention science [3], adult learning [4] and research utilisation. [24] As adult learners, policymakers are problem-centred and want practical ideas that have clear applicability [4]. They benefit from being involved in the content development. [4, 20, 28] Workplace learning benefits when presenters understand the learner in context and use appropriate narratives. [29] Vivid and detailed content increases learning, as does modeling struggle rather than effortless success. [30] Critical reflection — a well-established strength in individual and organisational learning — is positively associated with participation and self-efficacy (M5). [31] Data collection burdens can decrease willingness to participate overall. [32]

Mechanism 5. Self-efficacy

Self-efficacy refers to people’s beliefs about their capability to perform tasks and achieve goals. [28] Those who feel they have the skills to put ideas and resources into action are more likely to adopt them [2, 33], and more likely to welcome and commit to organisational change initiatives. [34] Thus self-efficacy increases motivation and persistence, but can also function as a vicious circle of self-doubt. [20, 35]

Interventions are more likely to build self-efficacy when they identify strengths, use credible experts to model values and behaviours, and enable participants to practice new skills. [28, 35] Participants should feel they are knowledgeable and valued partners in the learning process, and have a safe space for reflection (M7). [2, 4] Leaders play an important role in building self-efficacy via showing confidence in their staff and providing encouraging feedback. [35] Feedback can be used to create a sense of ‘wins’ in relation to intervention goals. [6] Existing beliefs about self- and organisational-efficacy affect the extent to which participants are willing to acknowledge scope for improvement and feel capable of owning the intervention goals—“If people cannot take care of a problem, they won’t see a problem”. [10:8]

Mechanism 6. Respect

Engagement and appreciation is a two-way process, participants need to feel valued in order to value in return. [36] Other studies emphasise the need for mutual respect in researchers/policymaker interactions [37] yet researcher/policymaker relationships are all too often characterised by separation, mistrust and poor understanding. [38, 39] Archetypal stereotypes of researchers and policymakers continue to have currency. [40]

Strengths-based practice advocated in social work and education, and more recently in public health, emphasises the importance of respecting participants via inclusion in intervention design and recognition of capabilities [41], while adult learning, social cognitive theory, audit and feedback theory and implementation science all emphasise the importance of participation and feedback on engagement. [2, 18, 28, 42]

Much of the more sophisticated research utilisation literature describes policymaking in terms of craft in which research is adapted rather than adopted [e.g. 24, 43], and implementation science makes a similar point in relation to interventions overall — they must be flexible enough to accommodate effective local practices. [e.g. 2, 3, 44, 45]

Disagreements can provide a platform for demonstrating goodwill and commitment to the relationship. [46] Efforts made by researchers to resolve concerns may increase the perception of being respected and instil greater confidence in their intentions.

Mechanism 7. Confidence

Frontline staff want initiatives to be credible and trustworthy, i.e. to have a strong evidence base but not threaten professional autonomy e.g. [27]. Thus potential participants weigh up “the quality and validity of evidence supporting the belief that the intervention will have desired outcomes”. [2] The source (internal or
external) and perceived legitimacy of an intervention affect how people engage with it. [47] People are less likely to embrace an intervention where it carries risks [3] This also applies to internal facilitators (M9) who may feel they are taking risks in championing the intervention. [48] Financial costs can affect attitudes towards an intervention [2] but it is not known how receiving free content affects confidence or perceptions of an intervention’s value in policy settings.

People who have endured multiple workplace change initiatives with little perceived benefit become suspicious and pessimistic. [6] Indeed, scepticism can be a major barrier to engagement, and often requires specific strategies to address it. [16] Attending to the history of an organisation is critical when developing appropriate drivers for future change. [49] Leaders can influence staff confidence, but may be dismissed if the intervention is seen as a top-down managerial initiative that is out-of-touch with on the ground realities, e.g. [10, 27].

The efficacy of audit and feedback is dependent on the perceived legitimacy of the data that informs it. [50] The quality and style of facilitation, including its sensitivity to context and support for local interpretation, is crucial for engaging potential participants. [51, 52] For example, feedback must be delivered constructively with an emphasis on continuous improvement and capacity rather than conveying connotations of surveillance or blame. [50, 53]

**Mechanism 8. Persuasive leadership**

As the ‘holders’ of an organisation’s values, leaders are key agents in effecting workplace change in using evidence in health [54] and in policy. [16] Without strong visible leadership and a persuasively communicated ‘vision’ for change, interventions will be stymied. [6] Leaders’ commitment to the intervention affects outcomes through multiple channels such as providing direction, shepherding the implementation and inspiring staff. [15]

Leadership can be particularly important in encouraging receptivity to ideas that challenge practice-as-usual [3], especially when they have a ‘transformational’ (facilitative) rather than ‘command and control’ style of leadership. [52] However, the power of leadership support can be overstated: leaders’ input is interpreted and is only one factor in many, so not even leaders’ best efforts will necessarily secure enthusiasm. [27]

Modelling refers to the way that people expand their knowledge and skills through observing others. This vicarious learning, particularly via leaders’ behaviours, conveys organisational values and norms. [20] The status and prestige of external experts who model wider values and practices will enhance their impact. [30]

**Mechanism 9. Strategic internal facilitation**

Not all facilitators will implement the intervention as planned [2], or to the necessary extent. [52] Thus selection of people with attitudinal commitment to, and appropriate skills for this role, is a vital consideration [55, 56], yet this is an often overlooked part of implementation. [2] Facilitators may require substantial internal and external support as they wrestle with “the compelling forces of fear of change, inertia, and investment in the status quo combine[d] with the inherently difficult and complex work of implementing something new”. [57]

In interventions, communication style, imagery and metaphor must be tailored for local participants [2], so effective marketing that taps into people’s different motivations requires inside knowledge. [45] To function as change agents, internal facilitators must be credible advocates of the intervention values they are espousing, and receive support from managers. [45, 56] Researchers and policymakers often benefit from help in bridging their communication divide [58], particularly as disciplinary paradigms run deep, thus facilitators of research utilisation interventions may need to play an especially nuanced role in managing expectations, negotiation and problem-solving. [59]
Appendix 15. references


43. Majone, G., Evidence, argument, and persuasion in the policy process. 1989, Yale University: Yale University Press.
Appendix 16. Publications arising from this thesis

Five papers have been produced as part of this thesis. Four have been accepted by peer-reviewed journals (three are published and one is in press), and the fifth has been submitted. They are listed below in the order in which they are presented in the thesis.


Contributions to other publications during my candidacy

Makkar, SR., A. Haynes, and A. Williamson, Health policymakers’ engagement with and use of research in the development of policies and programs. Under review with *PLoS ONE.*


Redman, S., T. Turner, H. Davies, A. Williamson, A. Haynes, et al., The SPIRIT action framework: a structured approach to selecting and testing strategies to increase the use of research in policy. *Social Science & Medicine,* 2015. 136-137, 147-55. [https://doi.org/10.1016/j.socscimed.2015.05.009](https://doi.org/10.1016/j.socscimed.2015.05.009)


Haynes, A, et al., Protocol for the process evaluation of a complex intervention designed to increase the use of research in health policy and program organisations (the SPIRIT study). *Implementation Science,* 2014. 9, 1-12. [https://doi.org/10.1186/s13012-014-0113-0](https://doi.org/10.1186/s13012-014-0113-0)

CIPHER Investigators, Supporting Policy In health with Research: an Intervention Trial (SPIRIT)—protocol for a stepped wedge trial. *BMJ Open,* 2014. 4. [http://dx.doi.org/10.1136/bmjopen-2014-005293](http://dx.doi.org/10.1136/bmjopen-2014-005293)

Oral conference presentations based on this work


Haynes A: Key findings from the process evaluation of a research utilisation intervention in six Australian health policy agencies. At Health Services Research Association of Australia and New Zealand (HSRAANZ) 9th Health Services and Policy Research Conference. Melbourne; 2016.


Invited blog

Appendix 17. Acknowledgements

I would like to thank the people who participated in SPIRIT and, especially, in my research. They gave their time generously to data collection and were both thoughtful and forthright in their responses – a qualitative researcher’s dream.

The process evaluation contributors listed in the authorship attributions section were a rock. I particularly want to thank them for their patience with me in the early fumbling stages of my research and when I requested feedback on drafts that were later changed beyond recognition. In addition to being a key member of this team, Sue Brennan used her spare time to immerse herself in the realist evaluation literature and joined me in wrestling with its application. She offered insight and much needed support when I had fallen down realist, quantitative and fidelity-related rabbit holes. She also laughed at my jokes. If karma exists, Sue is in for some good stuff down the track. Samantha Rowbotham offered similar help on the review paper, working with enviable efficiency and graciousness to unreasonably tight timeframes. Despite the challenges, this work was a pleasure. Thank you both!

Some of those who had important roles to play in the early days of this research were: Tari Turner, Nicola Lewis and Carmen Huckel-Schneider. Gai Moore was a PhD pioneer in my workplace and has been a valued buddy throughout, as has Mel Anderson. Sian Rudge and the rest of the smart, friendly and fun Knowledge Exchange and Evaluate teams make my workplace a nice place to be – they are hugely appreciated. Special thanks to Simone De Morgan for being a sweetie. Before all this, Michael Olsson at UTS introduced me to social constructionism and academic writing, and generally messed with my head. Importantly, he told me that, as an ex-social worker, I would make a good qualitative researcher. The seed was sown. Simon Chapman at the University of Sydney introduced me to health policy research, gave me the opportunity to stretch my wings as a researcher, and championed my further development.

Warm thanks to my supervisors. Anna Williamson’s expertise, steady presence and good humour made me feel that I could safely embark on this thesis. She provided wisdom, support and a calm space that I used to bounce around ideas and, on occasion, to mewl and rant. Phyllis Butow offered no-nonsense guidance that helped me identify which track I should be on, and provided direction when I couldn’t see any track at all. She was always responsive, even when juggling far more pressing demands. Sally Redman supported my work practically and conceptually. This included generous resourcing for the process evaluation and for my professional development. Despite different leanings in paradigm and methodology, and despite some early wobbles in my confidence, Sally demonstrated faith in me throughout. I have experienced misery, terror and delight as a result of this faith, but ultimately it had the effect that I suspect Sally intended – I strove to live up to it. I hope this thesis will be viewed as evidence that I have.

On a more personal level, love and thanks to Mike Price. Mikey, you are a star. I also recognise the contribution made by Jezebel and Charlie – there are mental health benefits to writing with a cat on one’s lap. Many thanks to family and friends who offered escape via movies, meals and conversation, and who accepted my ban on PhD inquiries with good grace; and to my mother, Jannettja, for being a formidable role model and believing that I would become a doctor in a matter of months.

Lastly, I could not have reached this point without the existence of the Australian Commonwealth Government’s Research Training Scheme which covered my tuition fees, or the financial support of two institutions: the NHMRC who funded the grant my research stems from and awarded me a postgraduate scholarship; and the University of Sydney’s School of Public Health which facilitated my PhD journey and granted annual scholarships that enabled me to take advantage of valuable opportunities to study and present my work overseas. Thank you.