Using formal theory in evaluation – What is it and how is it used?

Kim Grey
Charles Darwin University, Australia

Abstract
Little information is available about how formal, social science theory has been used in outcome evaluations. This gap exists in debates about what role theory has in evaluation. To help address this, we need to understand variation in formal theory use in real-world practice. This research applied systematic qualitative coding to identify and classify patterns of theory use, followed by qualitative comparative analysis. The terms describing theory are defined and formal theory is differentiated from other kinds of theory, such as program theory. The study found instances of borrowing and repurposing theoretical material in a cohesive sample of 17 outcome evaluation reports covering programs addressing complex social problems, drawn from cross-cultural contexts. Theory was mostly used for post hoc explanation and less often used upfront in framing, design and conduct of evaluation. The concrete approach in the literature of applying formal theory to measure and explain causal pathways to behavioural outcomes was found, as was layered use of a range of theoretical material. This article offers insights that may assist evaluators to undertaken more sophisticated and reflective approaches to using theory. Thinking about ways we use theory could form part of the tool-kit of techniques available to evaluation practice.

Corresponding author:
Kim Grey, Charles Darwin University, Ellengowan Drive, Brinkin, NT 0909, Australia.
Email: kimgrey@internode.on.net
Keywords
complexity, formal theory, practice, research about evaluation, social science theory, theory based evaluation

What we already know

- Very little is known about the use of formal theory in evaluation.
- We know that social science theory plays a role in evaluation, as theory can enhance explanation by drawing on existing knowledge to clarify why programs fail.

The original contribution the article makes to theory and/or practice

- This article is based on a study of use of formal theory in a set of evaluation reports.
- It offers insights which may assist evaluators to undertake more sophisticated and reflective approaches to using theory.

Introduction

Evaluation plays an important role in helping to understand rather than only measure the outcomes of programs. However, evaluation sometimes struggles to explain program effects when addressing persistent social problems (Leeuw & Vaessen, 2010). The search for adequate explanation of effectiveness and efficiency is a significant feature of outcome evaluation. The term ‘outcome evaluation’ refers to evaluations that describe the results of interventions, with the term ‘impact’ used to mean a study covering the attribution of cause and effect – see Schwandt (2015) and Stern et al. (2012). One way to enhance explanation is through the use of theory in order to draw on existing knowledge. While not all writers agree on the need for theory, Chen and Rossi (1983) argue that theory can clarify why programs fail and potentially help avoid this by anticipating weakness.

The idea that good theory is useful is attractive. This idea, commonly attributed to Kurt Lewin, dates back more than half a century. The aphorism ‘there is nothing as practical as a good theory’ has lasting resonance – Bedeian (2016) suggests that it comes from a reference by Kurt Lewin to a much older statement. Greenwald (2012) considers how theory contributes to the development of method. In evaluation, Pawson (2013) has emphasised the practical value of theory in building understanding of the
contingency and generalisability of evidence about program outcomes. Theory, it is argued, can help to improve our understanding of what works in what context, and whether an effective program might work in a different context. This suggests that theory has the potential to help expand the contribution of evaluation to addressing social problems.

In defining evaluation, Gullickson (2020) notes the need for credible systematic judgement. Credibility is commonly considered to draw from robust methods. However, it can also arise from systematic use of the knowledge base relevant to the evaluation questions, or what is already known about a social problem or the program or policy response.

While some authors make a case for theory in evaluation, Scriven (1998) argues that evaluation does not need to consider how or why programs work. Rather, he suggests that to develop the evaluation discipline, evaluators need some understanding of theory about evaluation to avoid errors of practice. Evaluators are not theoreticians as Gauthier et al. (2004) note, but evaluation is an emerging profession struggling with its’ identity. Wanzer (2021) notes the lack of an ‘evaluator’ identity makes it difficult for evaluators to gain status. Unlike research, social science theory is not embedded in evaluation. In her study of American evaluators and researchers’ perspectives, Wanzer (2021) found that only a small fraction (2.1%) of respondents reported that using social science theory is a part of evaluation. Instead, the purpose of generating knowledge for a client to inform decision-making and providing value judgements seem particular to evaluation.

There is a role for social science theory in the contribution of evaluation to knowledge generation. Gauthier et al. (2004) note two special characteristics of the evaluation discipline that can contribute to generating knowledge: the applied use of theoretical thinking (notably in program logic), and provision of defensible empirical evidence. The place of theory in evaluation is multifaceted and not yet settled – there is much to explore if we are to understand how theory is used and what benefits it offers. One benefit that more overt application of formal theory might bring to the evaluation profession is to help raise its academic credibility, a benefit of overt theoretical approaches that is flagged by Guenther et al. (2023).

Theory comes in many shapes and sizes. The meaning of words, such as theory, formal theory or program or policy theory and the concepts underlying such terms, are not always consistent or explained (Stern, 2023). For the purpose of this article, ‘formal theory’ is, broadly defined as the existing research-based theory that helps explain the relationships and patterns we observe in the world. Formal theory is distinct from program theory, and different from the concepts or propositions, frameworks and principles, which are also found in research literature.

Vaessen and Leeuw (2010) note some practical advantages of formal theory drawn from social science. They highlight the role that formal theories can play in anticipating challenges to program effectiveness. They add that formal theory offers a readily available and ‘cheap’ source of evidence, while also offering the potential to accumulate knowledge across interventions.
Evaluation practice draws many of its analytical methods from social science. However, there is a substantial gap world-wide between evaluation and the disciplinary knowledge concerned with understanding human behaviour and social systems (Donaldson, et al., 2009; Leeuw & Vaessen, 2010). One symptom of this gap is a lack of use of formal theory in evaluation. Stern et al. (2012) argue that disciplinary knowledge matters – which is the basis of formal theory – when we are interested in explaining, rather than merely measuring impact. Formal theory makes connections, which helps make sense of data and supports contextualisation and interpretation, and formal theory explains the links between causes and effects (Stern et al., 2012). Formal theory offers a bridge, allowing evaluation to reach into the knowledge base and obtain evidence to substantiate connections.

Stern et al. (2012) claim that use of knowledge or theories is particularly relevant to addressing complex social problems, such as those that are persistent, layered and multi-dimensional. Bridging the gap between evaluation and disciplinary knowledge by using existing theories can help to understand the interconnected effects of social programs. The validity of causal assumptions can be strengthened using existing knowledge. Krueger and Wright (2022) make a case that theory-based approaches (particularly program theory) complement approaches that explicitly address complexity, by addressing the challenges of measurement validity faced in ex-post outcome evaluations. When addressing complexity, we need every methodological tool available.

Several forms of theory-based evaluation have emerged as ways to respond to the need for evaluation of complex interventions (Brousselle & Buregeya, 2018). Contribution analysis, logic analysis and realist evaluation are examples of evaluation that draw on theory in several ways in order to help explain the mechanisms generating intended and unintended outcomes.

The lack of evidence about the use of formal theory (in particular) in evaluation practice is highlighted by Vaessen and Leeuw (2010). In the area of health behaviour interventions, the value of social science theory compared with contextualised theories of change is a subject of debate, due to the need to consider complex systems (Moore et al., 2019). There may well be limits to the transferability of formal theories to complex interventions operating in complex systems.

It could be argued that the case for seriously considering increased use of formal theory in evaluation itself is supported by the tendency of research about evaluation to draw on social science theory. Kupiec et al. (2023), for example, make use of organisational theories in their study of what organisational factors affect evaluation use. Linnell and Montrosse-Moorhead (2024) makes use of social identity theory in their study of evaluator identity. While building on existing knowledge is a foundational element of research, its place in evaluation is not so clear.

To explore this issue, this article draws on research into the potential for formal theory to make a stronger contribution to evaluation, particularly in outcome evaluations of programs addressing complex problems. This question is examined in a cohesive sample of 17 outcome evaluation reports covering programs that address
complex social problems or disadvantage, and which aim to assist Indigenous Aus-
tralians. Indigenous programs referred to here are social programs targeting the
complex needs of Indigenous Australians (also referred to as Aboriginal and Torres
Strait Islander people). This practitioner-led research focuses on understanding and
developing evaluation practice (Shaw & Lunt, 2018). While this research draws on
evaluations of Indigenous programs as a source for theory, exploring the use of theory is
intended to foster reflection about the issues more generally. This study does not relate
to the use of theory specifically in Indigenous contexts. The research draws on
evaluation reports relating to cross-cultural contexts, supporting reflection on ways that
evaluation reports appear to understand or make sense of the world. As Tovey and
Skolits (2022) indicate, evaluators don’t have much time for reflective practice. This
area of professional competency involves awareness of relevant content areas, as well
as awareness of one’s evaluation skills. In order to increase reflection about what
knowledge content evaluators use in their work, it is helpful to be more aware of how
evaluators use elements of knowledge, such as social science theory.

This article covers the role and meaning of theory in general and formal theory in
particular, and when and how formal theory might be used. The conceptual
framework and methodology used to examine the nature and use of formal theory in
a small area of evaluation practice are described. The method used systematic
coding of data drawn from document review along with comparative case analysis.
The findings draw from 17 outcome evaluation reports published over a six-year
period and limitations of this sample are noted. The findings may be relevant to
connecting evaluation and the disciplines on which evaluation needs to draw, if it is
to build on the stock of knowledge than can help to address social complexity. The
implications of this research for evaluation practice and guidance arising from the
findings is also provided.

Scope and terminology

Many definitions or descriptions of theory exist in the literature. ‘Formal theory’ can be
described as substantiated, research-based or verifiable theory that plays a role in
explaining relationships and patterns – their nature, order and connections – and the
reasons behind these relationships (Gay & Weaver, 2011; Babbie 1989). As a form of
applied research, evaluation can draw on a wide breadth of theoretical material. This
paper does not cover questions about program theory or evaluation theory, although
their use may be related to the use of formal theory. This is noted by Donaldson and
Lipsey (2006), who argue that there are three distinct kinds of theories important to
evaluation:

“…how evaluation should be practiced, explanatory frameworks for social phenomena
drawn from social science, and assumptions about how programs function and are
supposed to function.” (p. 57).
The roles of these three types of theory in evaluation – that is, evaluation theory, social science theory (or formal theory) and program theory – are distinct but overlapping. A common role played by all three kinds of theory is in communication. Making any theory overt helps communicate about it, defend and justify an approach and discuss the merit and worth of the evaluation itself.

There is limited guidance or research about formal theory or how it might be used. This study took a broad view of formal theory for two reasons. First, because it is not obvious how a reader might identify formal theory in an evaluation report. Second, a broad view is suitable in the context of research aimed at discovery, when little is known about use or the limits to usefulness. Similar to Jason et al. (2016), this is guided by the distinction between a context of discovery where findings are described, and a context of justification where predictions are tested. As there is no definitive database of all formal theory, a broad study allowed for the identification of as much material as possible.

Research design and methodology

The exploratory research questions addressed in this study into the real-world application of formal theories were:

1. What kinds of formal theories are used, in a sample of real-world practice?
2. How does their use vary?
3. What are the features of variation in the use of formal theory?

This research used systematic analysis and qualitative coding to describe and classify the use of theory. Examples of formal theory were identified and the strengths and weakness of applying the selected formal theories within the evaluation reports were examined. The methodology of this research considered aspects suggested by Walter (2010a), including standpoint, conceptual framework, and methods.

The focus on a body of material that represents real-life evaluation practice is informed by the authors standpoint as a practicing evaluator, exploring ways of conducting high-quality evaluation. The potential to inform practice guided choices made in this research design.

There are many challenges in the production of knowledge in a cross-cultural context. One challenge noted by Grey et al. (2018) in considering recent evaluation practice in the Australian context, is revealing an evaluator’s standpoint, to ‘make the invisible visible’ (p. 88). An underlying risk for evaluation is that value positions remain implicit.

Recognising potential bias is of critical importance in researching cross-cultural evaluation practice. While theory may hold universal potential to help make sense of reality, question assumptions or make predictions (Smith, 1999), legitimate concerns may arise about the relevance of the theories identified in this study to the cultural context of the reports in which they are located. Examining such concerns was not part
of the scope of this study. However, this study offers some insights into the diversity of practice in interactive and collaborative evaluation in cross-cultural settings, and the issues arising could be examined in further research.

**Sampling approach**

In order to begin to explore the topic of theory use, a sample of existing reports in a bounded policy area was used as the starting point to search for use of formal theory. This sample was a set of 17 existing outcome-focused evaluation reports, covering programs addressing the needs of Indigenous people. There is a lack of agreement regarding the definitions of outcome and impact evaluation (Schwandt, 2015). While definitions of ‘impact’ emphasis the attribution of any measurable difference to an intervention, definitions of ‘outcomes’ can simply infer that results are the product of an intervention’s effect (Schwandt, 2015). The reports were published between 2010 and 2015.

The research design involved a systematic document review complemented by comparative case analysis. The small size of the sample allowed coding of multiple features of reports and theories. This is a small population study similar to that by Miller and Campbell (2006) which is suited to descriptive data analysis, rather than tests of statistical significance. A descriptive approach is more appropriate to studies of how something is used in practice.

The research study accessed a database of evaluation reports established following the amalgamation of multiple Australian Government portfolios in 2013. The sample is drawn from this database of evaluation reports about Indigenous programs funded by the Australian Government, completed over a six-year period to 2015. It covers a broad range of policy topics: employment, education, crime, safety, child protection, welfare reform, environment, land, wellbeing, culture, governance and organisational capability.

From this sample, a sub-set of outcome evaluations was selected to examine evaluation practice that aims to diagnose policy or program lessons. The reports in the study sample that are referenced in this paper are marked in the references section (*). The full list of reports is available in Grey (2019). Evaluations that only examined process were excluded from the sample. The sample-set is cohesive in its coverage of programs that address complex social problems. Outcome evaluation in these circumstances confronts many challenges, particularly the adaptive or responsive nature of programs that address complex social problems.

Within the sample, two kinds of search strategies were used to locate theories. To identify any formal theories in this set of reports, citations were found and the source material was read to determine whether referencing pointed to any formal theories or other theoretical material. A wide framing was applied to avoid excluding items due to a lack of familiarity with a discipline. Word searches were undertaken to locate any formal theories that were not referenced. Initial search terms (i.e. theory, logic, model and standards) were used, with follow-up searches conducted to check for emerging
terms (i.e. principle and framework). Phase one of the research design involved identifying theoretical items. Phase two involved coding and analysis of the data.

**Coding and analysis procedures**

Document analysis was used to determine the frequency and type of theory found in reports. Systematic coding of documents based on a coding framework was applied. The coding process focussed on frequency, type and application of theory, and to capture any unexpected features. This involved multiple readings of the reports, based on processes similar to that described by Thomas (2006) as a ‘general inductive approach’ to qualitative data analysis. The characteristics of interest were summarised into descriptive coding matrices. The analysis contrasted cases to discern patterns and draw out strengths and weaknesses. Case comparison focused on issues that have the potential to inform practice. The results are drawn from an existing sample of evaluations of Indigenous programs, and analysis is not intended to provide generalizable

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axioms</td>
<td>Fundamental assertions</td>
</tr>
<tr>
<td>Concepts</td>
<td>Mental images representing and defining real things</td>
</tr>
<tr>
<td>Constructs</td>
<td>Theoretical creations, not directly observable but measurable (e.g. IQ)</td>
</tr>
<tr>
<td>Formal theory</td>
<td>Existing substantiated or verifiable theory (referenced and named) that helps understand and explain the relationships and patterns we might observe in the world</td>
</tr>
<tr>
<td>Frameworks</td>
<td>Parameters of mental models that structure, guide or limit the scope of what is important</td>
</tr>
<tr>
<td>Hypotheses</td>
<td>Specific, testable, statement about what is predicted to occur</td>
</tr>
<tr>
<td>Informal theory</td>
<td>Practitioner or evaluator theory, based on lived experience, past evaluation practice or constructed from other referenced material, that helps understand the world</td>
</tr>
<tr>
<td>Principles</td>
<td>Statements about what is so, summarising the way things are or the way things should be</td>
</tr>
<tr>
<td>Propositions</td>
<td>Statements expressing relationships between concepts, drawing on logical relationships between axioms</td>
</tr>
<tr>
<td>Statements</td>
<td>These include several kinds, for example, propositions, hypotheses, axioms and principles or laws</td>
</tr>
<tr>
<td>Theory</td>
<td>Systematic sets of interrelated statements intended to explain</td>
</tr>
<tr>
<td>Variables</td>
<td>Concepts made specific and measurable</td>
</tr>
</tbody>
</table>

This glossary draws on several references, that is, Babbie (1989), Gay and Weaver (2011), Jason et al. (2016), Leeuw and Vaessen (2010) and Walter (2010a).

Babbie (1989) equates principles with laws, or universal generalizations about a class of facts, while Patton (2018) describes two kinds of principles: ‘effectiveness principles’ draw from evidence (like the law of gravity), and ‘moral principles’ that are based on values.
conclusions, but to offer practice-based insights drawn from examination of use of theory in report writing.

Distinguishing between theoretical items is not straightforward. Where there was potential for overlap, for example, between formal theories and concepts or frameworks, two issues were considered: how the item was described in the report; and a check to establish whether the item is referred to as a theory in journal literature. The coding was based on the definitions in Table 1.

**Formal theory – operational definition**

A definition of formal theory is needed in order to study its use. In general, formal theory is research-based theory, which involves a robust and credible set of interconnected explanatory elements. Leeuw and Donaldson (2015) contrast two ways of thinking about theories on evaluation: research-based theories and practitioner-based theories (the latter are most commonly program theories and evaluation theories). Research-based theories are ‘Scientific theories capable of contextualizing and explaining the consequences of policies, programs and evaluators’ actions’ (p. 470). Formal theory is operationalised in this study as involving clearly conceived, interrelated explanatory sets of statements.

This study focusses on two of four elements of theories noted by Wacker (in Gay & Weaver, 2011). These primary elements are: distinct definitions of components; and inter-relationships between statements. Two further elements are domains and predictive claims. In this study, the four elements are treated as primary or foundational dimensions (definitions and relationships) and secondary dimensions, which build on the foundations by refining contingency and function (domain and predictive claims). While the secondary elements are relevant to understanding the potential functions of formal theory, the literature is less clear about the necessity of these elements. They are therefore treated as secondary dimensions, and are not applied as core features for the purpose of identifying formal theory.

**Findings**

This paper outlines the results of systematic coding and synthesis of the evidence contained in evaluation reports regarding use of theories. The research questions addressed what the kinds of theories found in this sample (question 1) and how did their use variation in application (question 2). This was followed by examining the features of variation in the use of formal theory (question 3). Part A describes the range of theoretical material found in the sample. It unpacks the characteristics of the types of material found. Part B explores the specific formal theories found in this sample, covering what kinds of theories they are and how they are applied. Comparative analysis was conducted in order to draw out the features of variation in these formal theories.
Part A — Range of theoretical material

This study examined use of theoretical material concerning people’s beliefs, behaviours and interactions within social systems, which is involved in assessing and understanding program effects. This material includes but is not limited to formal theory. Applying a wide conceptual frame means that the material identified in reports is considered in-scope if it is referenced to acknowledged sources.

The range of items. This sample of 17 evaluation reports contained more than 80 referenced theoretical statements (referred to as theoretical material or items). This material includes the ideas or elements of theories that are the building blocks of theories, as well as formal and informal theories. These items were used in 15 out of 17 reports.

These items of theoretical material are the subject of this section, rather than the evaluation reports. In order to answer the first research question in this study, the range of items of theoretical material found and their characteristics are described below.

Types of material. A wide range of theoretical material was found. The frequency of different types of referenced theoretical items is shown in Table 2. The largest group of items is propositions. Next most frequent are concepts and constructs. Formal theories make up around 10% of items (or eight items). There are also seven principles and six each of frameworks, hypothesis and informal theory. Axioms are least common. Table 2 shows that the majority of items are used to predict (46%) and explain (31%). Only one quarter of items were used to organise, categorise or describe. However, the majority of formal theories (5 out of 8 items) are used to explain.

Table 2. Types of items of theoretical material in use and their roles (counts and percentages).

<table>
<thead>
<tr>
<th>Types of items:</th>
<th>Organise</th>
<th>Categorise</th>
<th>Describe</th>
<th>Predict</th>
<th>Explain</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axiom</td>
<td>2</td>
<td>25%</td>
<td>11%</td>
<td>8%</td>
<td>12%</td>
<td>2 2%</td>
</tr>
<tr>
<td>Concept</td>
<td>1</td>
<td>25%</td>
<td>11%</td>
<td>8%</td>
<td>12%</td>
<td>9 11%</td>
</tr>
<tr>
<td>Construct</td>
<td>3</td>
<td>75%</td>
<td>14%</td>
<td>11%</td>
<td>23%</td>
<td>17 20%</td>
</tr>
<tr>
<td>Framework</td>
<td>1</td>
<td>25%</td>
<td>14%</td>
<td>11%</td>
<td>23%</td>
<td>6 23%</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>5</td>
<td>13%</td>
<td>11%</td>
<td>8%</td>
<td>7%</td>
<td>6 7%</td>
</tr>
<tr>
<td>Principle</td>
<td>2</td>
<td>22%</td>
<td>3%</td>
<td>4%</td>
<td>2%</td>
<td>7 8%</td>
</tr>
<tr>
<td>Proposition</td>
<td>2</td>
<td>43%</td>
<td>22%</td>
<td>31%</td>
<td>27%</td>
<td>4 28%</td>
</tr>
<tr>
<td>Theory -formal</td>
<td>3</td>
<td>8%</td>
<td>19%</td>
<td>5%</td>
<td>9%</td>
<td>8 9%</td>
</tr>
<tr>
<td>Theory -informal</td>
<td>2</td>
<td>5%</td>
<td>5%</td>
<td>2%</td>
<td>7%</td>
<td>6 7%</td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>85 100%</td>
</tr>
</tbody>
</table>

5% 8% 11% 46% 31% 100%

*aThis shows a count of items used. As one formal theory was used twice the number of formal theories is seven.*
This study revealed the use of referenced frameworks, principles and informal theories. Two kinds of informal theories were found: those drawn from referenced literature; and practitioner based (referenced to evaluators own prior work). Informal theories include operational theories focussing on program implementation and narrative descriptions of program impact theories, describing assumptions about how a program is expected to cause change. These were not called ‘program theories’ or ‘program logic’ in the reports.

**Variation in application.** The variety of use of theoretical material is described by why, where and how it is applied. This covers the purpose or role of theory in a report, application in the phases of evaluation, and the function that the theory fulfils.

**Why? Purpose of use.** This study explored the role of theoretical items in the report, categorised in three ways: as an independent reference source, justifying an approach or judgement, or iteratively, throughout the process of evaluating. The most common purpose seen in this study is that material is used as an independent reference, which potentially lends credibility to the results (43% of items are used in this way). Almost one quarter of the material is used to justify an approach or judgement. Over one quarter of items are applied in an iterative way, and revisited throughout the report.

Four of the eight applications of formal theories were used in all three ways in a report. In three reports, formal theories were not used iteratively – but were used to justify or explain a conclusion.

**Where? Phases of theory use.** In this sample, application of theoretical material was found to be spread throughout the phases of evaluation. These findings show that theoretical material was not used frequently in the earlier phase of evaluation – in study design, policy or program description, or in guiding measurement. Of the 85 items found, 40 items were used in single ways and 45 items were used in multiple ways. This study found many combinations of use:

- Only two items were used in all phases – that is, design, policy/program description, measurement, analysis and synthesis; both items were in *Department of Families, Housing, Community Services and Indigenous Affairs (2013)*
- Seven other items were used in the design phase – four in design alone and three in both design and analysis; these items were in four reports
- Nine items were used in a description of policy or program logic (with no other kind of use); (these items were in five reports)
- Nine items were used in program logic description combined with analysis and or synthesis (no items were used in both program logic and design) (five reports)
Seven items were shown in describing measurement alone, but these implicitly informed analysis (all items were constructs, used in one report (Harwood et al., 2013)).

Eight items were used in measurement combined with the analysis and/or synthesis phase – predominately constructs shown in reports by Australian Council for Educational Research (2011) and Department of Families, Housing, Community Services and Indigenous Affairs (2013).

A majority of the items (52 or 61%) were used in analysis and/or synthesis; there were 14 items used in analysis alone and six items were used in synthesis alone; the remainder were used in a combination of phases.

Half of the items used in analysis or synthesis were propositions and only one was a formal theory. Material was more frequent in analysis and synthesis than in any other phase, and this was seen in the multiple use of items.

Theory use was lowest in the evaluation design phase. In the small number of instances of use of theoretical material in evaluation design, over half of the items used were frameworks (4 out of 7).

The pattern of use of formal theory was diverse. One formal theory was used in all phases, including design – Social Influence Theory (Department of Families, Housing, Community Services and Indigenous Affairs, 2013). Two other formal theories were used in design. One evaluation report referenced a formal theory called Setting Theory (Harwood et al., 2013) and it was used to guide the evaluation design. None of the other six instances of formal theory use were in the design phase.

The six remaining instances of formal theory were used in the following ways. One report used Rational Choice Theory in program logic description only (Commonwealth of Australia, 2014). Two reports used formal theory in program logic description, analysis and synthesis – Collective Efficacy (Cooper et al., 2014) and the Capability Approach (Roche & Ensor, 2014). Two reports that used the same theory (Stages of Change Theory) used them in different phases. One report used this theory in analysis and synthesis (Colmar, 2014). Another report used it in program logic description and synthesis (Commonwealth of Australia, 2014). As well as using the Social Influence Theory in all phases, the report by Department of Families, Housing, Community Services and Indigenous Affairs (2013) used Social Identity Theory in measurement, analysis and synthesis. One feature of the instances of use of formal theory was their application in two or more phases.

_How? Functions of use._ The results of categorising how theory was used in each report show a strong pattern. Utilisation was unevenly spread across four categories of functional use of theory:

- **Framing** (formation of program theory and forming or prioritising evaluation questions)
- **Guiding** (scoping, planning, design and conduct of evaluation)
• Measuring (constructs postulated in referenced material)
• Explaining (causal effects, breakdowns, unintended outcomes).

This study found a strong emphasis on using theoretical material in explanation, and far less application in framing, guiding and measuring. Only 11 items (12% of all 85 items) were used in measurement, with four items used only in measurement and seven used in multiple ways. Three quarters of the items were used in explanation (63 items). The most common type of item used to explain findings were propositions. Concepts were also frequently used in this way, as were constructs; a large group of constructs were used for three functions (i.e. to frame, guide and explain). More than half of the formal theories were used to support explanation.

The majority of items were used once – 56 or 66% of all items, with most used to explain. One third of theoretical material was used in more than one way (29 items out of 85 items in total). The majority of multiple use was to ‘frame, guide and explain’ (21 items). Only a quarter of items were used in measuring along with other uses (framing and explaining: seven items). This illustrates iteration across phases of evaluation. With iteration, functionality appears to increase in analytical depth and move from abstract to more concrete application.

This coding shows a pattern of revisiting theoretical material for different purposes. However, this sample contains more examples of iteration that revisit ideas connecting descriptions of programs and analysis, than connections between framing, measurement and analysis. The iteration therefore remains abstract, as it is applied post hoc. This type of application is missing direct links between measurement and analysis.

**Summary of key features**

This study found variation in the use of a diverse range of theoretical material. Theoretical material has been applied in practice in ways that involves deeper types of use. This study shows low use of theoretical material in measurement, and low coverage of material in the action domain covering topics such as behaviour change or decision making (as described by Elster, 2007). This is despite many items drawing on the discipline of psychology.

Only a small number of formal theories were found. The smallest amount of theoretical material was found in the evaluation design phase.

Together, these findings suggest limited practical application of theoretical material to the prosaic purpose of measuring behaviour change. Instead, the use of theoretical material was seen more often in post hoc explanation, particularly in the use of propositions, in the analysis or synthesis phases of evaluations. Examples of this kind of use include unpacking breakdowns in causal pathways or implementation theory by supporting the questioning of assumptions. Further examination of the nature of use of theoretical material is provided in the next section, focussing on the use of formal theories.
Part B – Focussing on formal theories

The study located examples of formal theory by identifying theoretical material then narrowing the focus. ‘Formal’ theory is a sub-population of the items of theoretical material found in this study. This aims to understand the range of formal theories and how they are used, based on comparative analysis. The general function of each theory is identified, along with the specific nature of its use in context. This comparative case analysis supports the identification of strengths and limitations of the use of formal theories in new contexts. Significant features suggested by the comparisons are listed at the end of this section.

Range of formal theories. The key focus of this study is to examine the use of formal theories in evaluation. Eight instances of the application of formal theories were found within six reports. One report applied two different formal theories; and one report applied two formal theories. Given that two reports used the same theory, seven formal theories were found in the sample. These are grouped by the roles they play, showing the categories of understanding to which they contribute (i.e. explanatory, contextualising and process). The seven formal theories are shown in Table 3.

Six formal theories found in this study draw from three disciplines: three formal theories have roots in psychology, one in sociology, and two in economics. The seventh formal theory concerns research itself. Most are in the action domain and most are explanatory.

All seven theories are described in Table 4 below. Each theory is described briefly in the next section, followed by a short analysis of the application of each formal theory. Two examples are explored in more depth, as they offer insights into the application of formal theories based on comparison.

The guiding definition of formal theory used in this study is ‘research-based theory’ that involves a robust, clear and credible set of interconnected explanatory statements. The formal theories found in this study cover explanation as well as contextualisation. These theories help understand the consequences of interventions.

The first two theories shown in Table 4 (examples 1 and 2) are drawn from social psychology and relate to how social context drives behaviour. The next three formal theories

<table>
<thead>
<tr>
<th>Table 3. Formal theories found by type of understanding to which each theory contributes (N = 7).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explanatory (n = 4)</td>
</tr>
<tr>
<td>Social influence theory</td>
</tr>
<tr>
<td>Social identity theory</td>
</tr>
<tr>
<td>Rational choice theory</td>
</tr>
<tr>
<td>Stages of change theory</td>
</tr>
</tbody>
</table>

*aThe categories of type of understanding shown here draw on Leeuw and Donaldson (2015).*
Theories (examples 3, 4 and 5) help understand the nature of the world (i.e. covering the relevant context or processes) and support research into the phenomenon of interest. This includes a social change theory, a critical development theory, and a theory about methodology. The common features of these three theories are the use of interrelated statements about distinct concepts to explain foundational or background phenomenon, which supports the contextual understanding of evaluands. The last two theories in Table 4 (examples 6 and 7) are behaviour change theories that offer explanation of factors or states related to the individual decisions that guide human actions.

Research-based formal theories found in this sample are all well-conceived, referenced, interrelated explanatory sets of statements that support our understanding of what shapes social behaviour. They vary, however, in the extent to which they predict

---

**Table 4. Description and domain of all formal theories found in this study.**

<table>
<thead>
<tr>
<th>Formal theory</th>
<th>Description (domain)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example 1: Social influence theory (Kelman, 1958)</td>
<td>Social psychology theory about how behavioural changes are influenced through social processes in three ways: compliance, identification and internalization (action domain).</td>
</tr>
<tr>
<td>Example 2: Social identity theory (attributed to Henri Tajfel in McLeod, 2008)</td>
<td>Social psychology theory about how a person’s concept of self depends in part on the groups that provide our sense of belonging to the social world (McLeod, 2008) (interaction domain).</td>
</tr>
<tr>
<td>Example 3: Collective efficacy (Sampson et al., 1997)</td>
<td>A social change theory about how violence may be reduced in the context of greater social cohesion and trust, involving community willingness to intervene on behalf of the common good (situated in neighbourhood context) (domain of the mind).</td>
</tr>
<tr>
<td>Example 4: The capability approach (attributed to Amartya Sen) (Robeyns, 2005)</td>
<td>A critical development theory that identifies the role of human capabilities to function, (what people are able to be and to do), in supporting wellbeing and development (action domain).</td>
</tr>
<tr>
<td>Example 5: Setting theory (Barker, 1968; Cambourne &amp; Kiggins, 2004)</td>
<td>Methodological theory based on ecological psychology, about the relationship between patterns of behaviour and features of context (interaction domain).</td>
</tr>
<tr>
<td>Example 6: Rational choice theory (Australian Public Service Commission, 2007)</td>
<td>Theory about behavioural choices covering how actions or decisions are based on rational assessment of the costs and benefits of available choices (action domain).</td>
</tr>
<tr>
<td>Example 7: Stages of change theory (Prochaska &amp; DiClemente, 1983)</td>
<td>Behaviour change theory about stages of people’s motivational readiness for action and how these stages match with processes of change (action domain).</td>
</tr>
</tbody>
</table>
(as well as explain) possible outcomes, and in the nature of outcomes or ‘states of being’ that they help explain. Therefore, some theories offer more contextual understanding of elements of the phenomenon of interest in these evaluations (rather than predicting behavioural outcomes).

Summarising the nature of the seven theories shows this variety. Two theories offer explanations of how behaviour change or decisions are predicted to occur (i.e. Rational Choice Theory and Social Influence Theory). The combination of factors that influence an effect are outlined by the Collective Efficacy theory. This is the only formal theory found in this study that is drawn from a referenced study situated in the same social problem as that being addressed in the report in this sample – an evaluation of crime prevention by Cooper et al. (2014). Setting Theory is also unique in this sample as it guides the conduct of research, explaining how setting influences the data collected by evaluators.

Other theories help explain the contextual phenomena that influence the operation of an evaluand. The Capability Approach explains a relationship between states of being and doing by offering connected concepts that support contextual understanding of factors that are necessary to development. The Capability Approach, which is referred to as both a theory and a framework, can be applied at different levels of abstraction – showing how a theory can become an input into another theory, or model, at a different level of abstraction. The Social Identity Theory explains how our sense-of-self is affected by group belonging, informing the operation of Social Influence Theory. This ‘layering’ of theories is highlighted in the application of the Stages of Change Theory which is both a theory in its own right (predictive of people’s readiness for change) and a construct or factor in a wider impact model.

The application of all seven formal theories found in this study is summarised in Table 5, showing use of formal theories in single and multiple ways. Of the four formal theories in the action domain (examples 1, 4, 6 and 7), three were used iteratively.

**Summary of results**

This study provides insight into the range of uses of theoretical material, drawing on a sample of recent evaluations of programs that address complex, persistent social problems or multi-dimensional disadvantage. The research found more instances of borrowing and repurposing theoretical material in post hoc explanation. Less common was the use of theoretical material upfront in framing an evaluation or iterative use of theories throughout the design and conduct of evaluations. In summary, the three core findings are as follows:

- A wide range of theoretical material, beyond formal theory, was used in these reports, much of it for explanation, with less found in the earlier phases of evaluation design
- There was limited use of theoretical material in measuring or explaining action or behaviour, and more explanation related to mental constructs or social connections or relationships (i.e. the domains of the mind or interaction)
The functional use of ‘theories’ in this sample varies along a spectrum of increasing analytical depth, involving more iterative and reflective application.

The study also found two further features of variation. These are drawn from comparative analysis of variation in the use of theoretical material:

- Reports repurpose formal theories, plus a wide range of theoretical items (the potential component parts that make up the building blocks of theories), as well as wider conceptual resources
- The overlapping roles of theoretical material: it is not just formal theory that is used in these reports – other kinds of theoretical material also fill similar kinds of roles (i.e. in framing, guiding, measuring and explaining).

Evaluation reports were found to draw from disciplinary knowledge in a wide range of ways. This study aimed to examine the use of formal theory in particular. However, by

<table>
<thead>
<tr>
<th>Formal theory</th>
<th>Categories of use</th>
</tr>
</thead>
</table>
| Example 1: Social influence theory | Used iteratively (Department of Families, Housing, Community Services and Indigenous Affairs, 2013):
- To frame the design of a welfare reform evaluation
- To measure and explain signs of behaviour change. |
| Example 2: Social identity theory | Informs measurement and explanatory analysis in a welfare reform evaluation (Department of Families, Housing, Community Services and Indigenous Affairs, 2013); used in combination with Social influence theory. |
| Example 3: Collective efficacy | Used to explain potential areas of theory failure in an Indigenous justice program evaluation of crime prevention (Cooper et al., 2014). |
| Example 4: The capability approach | Used iteratively in an evaluation of community development (Roche & Ensor, 2014):
- To frame and guide a theory-based evaluation design
- Help explain findings and assess effectiveness. |
| Example 5: Setting theory | Guide evaluation design and conduct of data collection in an evaluation of an education mentoring program (Harwood et al., 2013). |
| Example 6: Rational choice theory | Used to frame the assumptions in a theory of change in an evaluation of school attendance (Commonwealth of Australia, 2014). |
| Example 7: Stages of change theory | Repeated use in two contexts:
- Iteratively to explain breakdowns in causal models predicted by a theory of change in an evaluation of school attendance (Commonwealth of Australia, 2014)
- To explain implementation failures in an evaluation of a restorative justice project (Colmar, 2014). |
adopting a broad conceptual framing, the evidence revealed features of use of theoretical material that were not originally envisaged. Adopting a wide conceptual frame was intended to help avoid bias. This wider frame challenges the boundaries around what kind of ‘theory’ can be useful in evaluation practice. It supports the examination of descriptive, organising and normative theoretical and conceptual material, rather than focusing exclusively on the explanatory and predictive functions of theory. An additional benefit of stepping out from a narrow framing is that it allows observation of how formal theories are used, in contrast with other forms of theoretical material.

**Implications**

This research provides some insights which may have implications for evaluation practice. Guidance arising from the findings is provided, covering ways of using social science theory in evaluation practice. It should be noted that this is an exploratory study which did not set out to determine whether using theory makes evaluation more effective. It was not designed to provide definitive guidance. Instead, research conducted for the purpose of discovery aims to help understand a phenomenon more clearly. This research set out to explore, in a modest way, how theory is used and how we might grow our understanding about its use.

Several implications of this research for evaluation practice are offered:

- Evaluators could be more reflective about what theories we choose to use and why we choose to use them in the way that we do.
- Evaluators could be more overt and transparent by documenting our choices.
- We could consider if there are other theories that we could use, beyond those we are familiar with.
- We could ask whether the theoretical material chosen is sufficient.

These implications call on us to raise our awareness of the reasons behind our choices. It suggests that we could consider whether the selection of theories is based on sufficient reflection about the theoretical material that could potentially be used. Use of theory in evaluation can be wider than use of formal theory or program theory. It may involve use of a range of theoretical materials or conceptual resources, drawn from disciplinary knowledge.

More specifically, the limited use of formal theory found in this study in the design or measurement stages of evaluation could limit the clarity and transparency of evaluation design or findings. For example, what are the risks if the underlying assumptions implicit in the selection of constructs chosen for measurement are not expressed? Could this reinforce conventional thinking? Being overt about the mental models and sources of knowledge that underlie our choices means being more aware about the standpoint that frames choices in evaluation design.

The low level of iterative use of theories across phases of evaluation, such as in connecting conceptual framing, measurement and analysis, suggests post hoc selection
of theories. Using theory may help explain findings by reference back to related literature. Revisiting theoretical material during phases of the evaluation may also be useful in explanation. In doing this, evaluators should consider the potential bias that could arise from the lack of upfront links between use of theory in measurement and in later analysis. Single use of formal theories to justify or explain a conclusion may at best miss an opportunity for deeper analysis. At worse, it may reinforce bias.

There are also implications for our thinking about systems. We could consider whether the theoretical material chosen is sufficient, that is, whether it addresses the layers of complex systems in which the intervention operates, or the layers of the intervention itself. For example, is the theory chosen related only to one level of a system, such focussing on individual behaviour? We could consider if the choice to look at only one part of the system is justified in the circumstances of the particular evaluation.

In summary, the main points of guidance for using formal theory in evaluation practice offered to readers relate to the importance of being more overt in use of formal theory. A key ingredient in evaluation practice that is ethically accountable to the people it is ultimately intended to benefit, is our capacity to undertake reflective practice and question underlying or invisible assumptions. More reflective approaches are especially important to complex social issues and cross-cultural evaluation, suggesting a need to make evaluation design choices more overt. In practical terms, evaluation reports should document the choices made about using theoretical material and the ways in which theories are applied.

Future research into the use of formal theories could examine the pattern of results in other policy areas. However, research could also look further into different kinds of usage, such as the use of formal theory in building program theory.

**Conclusion**

This study has focused on the use of formal theory in evaluation practice. The research has explored three questions: what kinds of theories have been used; how use of formal theory varies in practice; and what features of variation might be important in understanding use of formal theory? An exploratory approach widened the frame of reference beyond formal theory to examine how broader, research-based, theoretical material might contribute to bridging the gap between disciplinary knowledge and evaluation practice.

‘Theory’ is a broad term, which is ambiguously applied. Many kinds of theoretical material are used in similar ways to formal theory. Evaluators borrow and repurpose formal theories, as well as a wide range of the component parts that make up the building blocks of theories, as well as wider conceptual resources. The lack of transparency about how evaluation is done and the choices made by evaluators could be holding back development of the field of evaluation. This study shows the multiple range of ways that evaluators reference, reuse and recycle theory in producing evaluation reports and suggests there is value in being more reflective about the use of theory.
Declaration of conflicting interests

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

Funding

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

ORCID iD

Kim Grey  https://orcid.org/0000-0003-4655-9339

Note: evaluation reports from the sample used for this study are marked (*).

References

Note: evaluation reports from the sample used for this study are marked (*).


