How Evaluation Affects Accountability Mechanisms

Author: Stephen Porter

Submitted for PhD by Publication

University of East Anglia, School of International Development

Submitted: July 2020

Abstract

In international development people with power often publicly demand accountability on the basis that it supports enhanced poverty reduction. Accountability mechanisms are forums where these demands for accountability are realised through agents meeting to exercise and constrain power. In exercising and constraining power, the agents in the accountability mechanism draw on evaluation to assist decision-making. Evaluators are familiar with providing evaluation as an instrument that responds to demands of accountability mechanisms (Christie and Alkin, 2014; Rossi et al, 2018). For evaluators though, there is a gap in research related to how they affect accountability mechanisms to support the achievement of a goal – be it responsiveness to the voices of marginalized people, social betterment, evidence use, or transformation. To close the evidence gap, this thesis presents a middle-range theory of how evaluation affects accountability based on a thorough review of relevant studies and a meta-ethnographic synthesis of five studies that I authored or co-authored.

The thesis proposes a middle-range theory for how evaluation affects accountability mechanisms through a strategic approach in which evaluators with other agents target achievement of a common goal, such as, expanded voice of citizens. Strategically the middle range theory prioritizes navigating the authorising environment in order to develop partnerships that provide leadership and exercise power towards a common goal. Three tactics implemented in the process of evaluation support strategic change in the authorising environment, namely, (i) expanding the focus of demand, (ii) accessing the agenda of agents; and (iii) undertaking a shared journey on evaluation quality. When evaluators work towards a common goal in an accountability mechanism's authorising environment and activate the three tactics in the process of evaluation, the logic of decision-making adapts.

In working to affect accountability mechanisms evaluators require relational and analytical approaches to engage in political spaces. This thesis identifies five approaches that can be combined to assist in both the authorising environment and evaluation processes. Evaluators need approaches to work in a relational manner, which in this study meant undertaking Helping (Porter, 2011; Schein, 2009) and joint production (Packard Foundation, 2010b;Porter, 2013). Evaluators also require approaches to map opportunities and constraints for affecting accountability mechanisms by, for example, by detailing the demand for evaluation (Porter and Goldman, 2013) and the political economy of the context (Porter, 2017). The Capability Approach (Alkire, 2005; Sen, 1999) arises in this study as a guiding framework for affecting accountability mechanisms. The capability approach goes further than the other approaches discussed in the study by providing an underlying ethic, which has both relational and analytical applications. The capability approach also has an in-depth justification for the prioritisation of freedom in both the means and ends of development, and an emphasis on a broad information basis for evaluation (Porter and de Wet, 2009).

Table of Contents

Tables

Figures

Acknowledgements

This thesis is dedicated to Daniel Plaatjies whose tremendous energy, commitment to improving the working of the state and kind mentoring have shaped my career and approach to evaluation. In completing this thesis, I have benefitted from support and guidance from colleagues, mentors and my family. Thanks go to the supervisors for this thesis Laura Camfield and Maren Duvendack who have guided me through the process and always been on hand to provide encouragement and advice. I have benefited during my endeavours to complete a PhD on evaluation from previous supervisors Patricia Rogers, Elizabeth Branigan and Ivor Sarakinsky and mentoring from Donna Podems, Bob Picciotto, Osvaldo Feinstein, Rodney Hopson, Andy Rowe, David Rider Smith, Ray Rist, Alison Evans, Thomas Mogale and Peter Kingstone. The collaborations and thinking developed with co-authors Penny Hawkins, Jacques de Wet and Ian Goldman flow throughout this study. Finally, many thanks and much love to my family Chevon, Isabella, Callen, Peter, Jane, Richard, Rachel, Jack and Tommy.

1. Introduction

International development articulates the centralisation of values of respect for human rights and human dignity, inclusion, the rule of law, justice, equality and non-discrimination (UN General Assembly, 2015). Yet, evidence shows that these values are often displaced by hypocrisy, social structures that reinforce colonialism and underdevelopment, and agents who ignore evidence which does not fit with their ideology (Ake, 2001; Escobar, 2011; Moyo, 2009; Stiglitz, 2002; Weaver, 2008). The definition and function of accountability mechanisms and evaluation practice are at the front-line of this gap between intention and action in international development.

Major International development agreements seek to end poverty and formally establish links between accountability mechanisms and evaluation to better align intended values, decision-making and results within national systems. The Paris Declaration links evaluation and accountability to work towards common principles for aid effectiveness guided by national ownership of development priorities (OECD, 2005). The Accra Agenda for Action was informed by an evaluation that found progress towards the Paris Declaration commitments was too slow, with agreements revisited and deepened to enhance aid effectiveness, including those related to country-led evaluations (OECD, 2008). The Sustainable Development Goals (SDGs), the most recent overarching global agreement to end poverty, states that the accountability mechanism of follow-up and review processes will be informed by country-led evaluations (UN General Assembly, 2015).

These agreements consistently cement the principle that people living in poverty should direct action through their voices[1] and engage in deliberation about the goals and results of development (Fox, 2015; Manroth et al., 2014; OECD, 2005, 2008; UN General Assembly, 2015). Evidence has also identified that voice and deliberation are key aspects in enhancing the effectiveness of development. For example, Amartya Sen (1999; 2005, 2007; 2009) has compellingly argued that engaging the voice of marginalized people, public reasoning and deliberation generate important conditions to alleviate poverty in drawing together research from economics, history, philosophy, political science and studies of culture, among other areas. In *Development as Freedom,* Sen (1999) provides evidence which demonstrates public discussion expands the opportunities for development and prevents serious deprivation, such as famine. These ideas are then further developed in Sen's book *The Idea of Justice* (2009, pp. 321-354) where he elucidates how the practice of democracy as "government by discussion" supports the expansion of social welfare, securing protections against deprivation and enabling minority rights and inclusive priorities.

[1] The terms 'voice' and 'voices' are often used to signify that people have meaningful input to policy processes that affect their lives (see for example Johnson and Rasulova, 2016; Fox, 2015; Hopson and Cram, 2018). When used in this thesis it intends no implication that a single homogenous voice of people exists. Sorting through the heterogeneity of perspectives represents a key role of accountability mechanisms and evaluation practice.

The argument presented in this study draws upon my seventeen years of experience in working in developing countries, fourteen of these being spent working on implementing monitoring and evaluating systems across international development institutions. In my career I have worked with the private sector, grassroots community service organisations, in international NGOs, bilateral donors, academia and multilateral organisations. The movement between these types of organisations was intentional as I found myself grappling with how to orientate development practice towards the perspectives of people living in poverty and realise the ambitions of the international agreements for country-led development. In each setting I have found practice constrained by organizational and institutionally imposed rules and cultures each with their own sets of opportunities. I discuss this experience and how it has framed my own positions in the next chapter.

1.1 The Connection between Evaluation and Accountability

When functioning well the link between accountability and evaluation mechanisms in international development should enhance efforts to reduce poverty in accordance with agreed values and do so in a manner that forefronts poor peoples' voices and reasoning. Accountability mechanisms should orientate their composition and their decision-making structure to align to these intended values and results. Though accountability mechanisms are constrained by structures that point decisions towards the interests of those with more power (Crush, 1995; Porter and de Wet, 2009), space can be opened to ensure that chains of agents engage with people living in poverty. Evaluation can assist in opening spaces by providing feedback that develops and summarises results and practice. Further evaluation practice informs how accountability mechanisms can better function through systematic processes to review evidence focused on values.

Though evaluation and accountability are intertwined concepts and practices, the linkages between the two have not been fully developed in theory. Accountability mechanisms are forums where agents meet to exercise and constrain power through reporting based on a set of criteria. Accountability mechanisms inherently engage in evaluation through enabling relations, providing answers, looking backwards and making decisions with potential sanctions. In exercising and constraining power, the agents in the accountability mechanism can draw on evaluation to assist decision-making. At the level of definition evaluation identifies its role as "a key analytical process in all disciplined and practical endeavours", that focuses upon "determining the merit, worth and value of things" (Scriven, 1991, p. 1). Evaluation needs to engage accountability mechanisms to inform action. Currently, evaluation orientates itself as an instrument for consideration in accountability mechanisms. Yet given the interplay evaluation has a potential role beyond an instrument by substantively affecting the way performance, partnerships, results and future action are considered by an accountability mechanism, even when not receiving a formal evaluation report.

The evaluation discipline has yet to build out theory that explains the substantive interplay with the structure of decision-making in accountability mechanisms. In some cases, evaluation talks around accountability, but does not use the word (e.g. Pawson and Tilley, 1997). In other

instances the purpose of evaluation is argued to be an instrument for accountability, amongst other purposes, such as learning, program improvement and knowledge generation (e.g. Chelimsky, 2006; Chilisa et al., 2015; Greene, 2006; Patton, 2008; Rossi et al., 2018). In some literature the debate focuses on rivalry between accountability and learning (e.g. Guijt, 2010; Kachur et al., 2016; Kogen, 2018; Picciotto, 2018). Concerns with providing evidence to inform changes in practice have often identified tactics to change evidence gathering processes to better consider power (Bell and Aggleton, 2016; Cornwall and Aghajanian, 2017; Eyben et al., 2015; Hayman et al., 2016; Hedrick, 1988; Johnson and Rasulova, 2017; Weiss, 1993) with some identifying elements of a more strategic approach (Goldman and Pabari, 2020; Hopson and Cram, 2018). A limited number of books and articles also envisage evaluation affecting the way accountability mechanisms function, but these ideas are not built out into theory (e.g. Fetterman and Wandersman, 2005; Mackay, 2007; Picciotto, 1995; Ryan, 2004, 2005; Van Der Meer and Edelenbos, 2006; Wiesner, 2011). As a result important pathways through which evaluation exercises and mediates authority remain partially developed; a middle-range theory can add value to this space.

This thesis contributes to closing a gap in knowledge on how the practice of evaluation affects the operation of accountability, specifically focused on international development. The thesis adds value to current debates by proposing an empirically grounded middle-range theory. Drawing on Fox's (2015) distinctions the theory anticipates that evaluation affects accountability mechanisms through a strategic approach to power relations in which evaluators with other agents target achievement of a common goal, such as, expanded voice and public reasoning. The middle range theory prioritizes strategically navigating the authorising environment in order to develop partnerships that provide leadership and exercise power towards a common goal. Meanwhile three tactics implemented in the process of evaluation support change in the authorising environment, namely, (i) expanding the focus of demand, (ii) accessing the agenda of agents; and (iii) undertaking a shared journey on evaluation quality. When evaluators work with others towards a common goal in an authorising environment and activate the three tactics in the process of evaluation, accountability mechanisms adapt their decision-making criteria and processes.

Evaluation practice has an important role to play in international development practice as it has enormous reach and influence in helping to assess the achievement of international agreements concerned with reducing poverty and inequality, such as the SDGs and the Paris Declaration and Accra Agenda for Action on Aid Effectiveness. Development evaluation operates in a multitude of contexts where institutions and lines of accountability between citizen and delivery are still forming within emergent, complex and sometimes chaotic structures (Reeler et al., 2009).

1.2 Developing Middle-range Theory Building on Experience

My professional experience in implementing monitoring and evaluation systems has been expressed in peer reviewed studies, six of which substantively inform this thesis. The studies incorporate around ten years of academic and practitioner experience. The six studies originate from undertaking evaluation in the field of international development, five of these studies are

further incorporated into the meta-ethnography that supports the development of the middle range theory (Porter, 2011; Porter, 2013; Porter, 2017; Porter and de Wet, 2009; Porter and Goldman, 2013), while one is referenced as a previous synthesis of experience (Porter and Hawkins, 2019).

Throughout my professional and academic work, I have often found myself drawing on the scholarship of Amartya Sen, which provides a clear goal for development and an ethical approach for undertaking practice. His expression of the capability approach I have found to have value at all levels of the development system, whether discussing evaluation systems with a Chief Economist or opportunities for enhanced self-care with community care workers. Upon writing this thesis I found myself identifying how Sen's work has informed my own by providing a guiding set of principles that centralises freedom and information breadth and recognizes that peoples should be the authors of their own destiny. Though not always referenced in my peer reviewed works the themes emanating from Sen's thinking remain consistent.

Drawing on studies that originate from dynamic international development contexts and that I am intimately familiar with as author and co-author provides insight on how evaluators are active in accountability mechanisms that are forming and reforming. The international development context stands in contrast to evaluation literature that, for example, originates from the United States, where established accountability mechanisms operate. Though the middle-range theory represented here originates from an international development context, it can be tested in any context, particularly those where accountability mechanisms are in flux.

Middle-range theory receives interest amongst evaluators because of its potential to produce explanations that can be carried between programs and policies. Middle-range theory focuses on developing a specific hypothesis from empirical investigation, consolidated into wider networks of theory, which are sufficiently abstract to deal with different spheres of social behaviour and structure (Merton, 1968; Pawson, 2000). Evaluation practice currently focuses on realist evaluation approaches to produce middle-range theory.[2] This study provides an example of middle-range theory focusing on evaluation practice itself and employing interpretivist, rather than realist techniques.

To develop the middle-range theory outlined, the study applied a meta-ethnography approach. Noblit and Hare (1988, p. 12) developed meta-ethnography as an approach to synthesise interpretive research that pursues "an explanation for social and cultural events based upon the experiences of the people being studied." Meta-ethnography originates in seeking explanation of broader changes in society and power dynamics and by synthesising five ethnographic studies of school desegregation in the United States in the late 1970s (Noblit and Hare, 1988). Meta-ethnography locates itself within the interpretive paradigm [3],placing emphasis on inference and explanation situated within multiple perspectives.

[2] See for example, the studies funded by the Centre of Excellence for Development Impact and Learning (2020)

[3] Noblit and Hare (1988) base their understanding of the term *paradigm* in the writing of Kuhn (2012) . Kuhn (2012) argues that a paradigm originates in scientific revolutions that break with previous perspectives, reshape

A meta-ethnography applies an approach to synthesise existing studies through seven phases that entail reading, juxtaposition, interpretation and synthesis. For this thesis additional guidance was drawn from France et al (2019), who undertook a methodological systematic review of meta-ethnographic studies. Additionally, meta-ethnographic studies and reviews that provide advice on working from a purposive selection of studies were utilised (Campbell et al., 2011; Campbell et al., 2003; Pilkington, 2018). Identification and interpretation of concepts in the meta-ethnography was supplemented by a focused review of books and articles that connect evaluation and accountability. The literature review entailed a search of nine leading journals of evaluation, snowballing of references, taking the advice of other experts and review of literature drawn from my own articles.

The thesis broadly follows the approach of meta-ethnography to represent the development of thinking and findings. This introduction is followed by a discussion of positionality, justification and outline of the meta-ethnography approach (chapter 2). The thesis then follows the seven phases of a meta-ethnography. Phase 1 defines the area of interest for meta-ethnography through presentation of a literature review that demonstrates how evaluation and accountability are currently connected in the literature and identifying gaps that are present (chapter 3). Phase 2 defines the scope of the research through stating the research problem, objectives, question and outlining the chapters and articles to undergo a meta-ethnography approach (chapter 4). Presentation of phases 3-6 of the meta-ethnography takes place in a single section, which also describes the methodology for translation, interpretation and synthesis (chapter 5). Phase 7 of the meta-ethnography expresses the synthesis and expresses the middle-range theory (chapter 6). Having presented the meta-ethnography a brief conclusion summarises arguments and highlights areas of follow-up (chapter 7).

understanding of previous knowledge and provide a new set of rules and standards guiding scientific enquiry that become assumptions by those who share the new rules.

2. Justification: The Meta-ethnography Approach

The meta-ethnography approach applied in this thesis draws on broader qualitative practices that seek to synthesise data and translate themes and so generate new concepts (Gehman et al., 2018; Hancock and Algozzine, 2017; Noblit and Hare, 1988; Patton, 2014). To synthesise 14 years of professional and academic practice in monitoring and evaluation required the selection of an approach that would synthesise text and accommodate contextual reflections not fully apparent within individual peer reviewed publications. The meta-ethnography approach enables the synthesis of qualitative data and maintains contextual elements to generate new insights and concepts (Campbell et al., 2011; France et al., 2019; Noblit and Hare, 1988). This chapter, firstly, provides an examination of my own beliefs, judgements and positions feeding into this study. Second, a justification is provided for why meta-ethnography helps to usefully synthesise my practitioner and academic experience with rigor. Third, the seven phases of meta-ethnography are discussed in detail below, where the penultimate phase of translation stands in contrast to techniques where data is aggregated with the aim of knowledge accumulation (see for example, Duflo and Banerjee, 2017; Eisenhardt, 1989).

2.1 Positionality

My approach to evaluation has been influenced by the lessons garnered and imparted during a professional and personal journey across the levels of the aid and development system. From this journey two main lessons emerge that are pertinent for this thesis. First, the values of international development exist in tension with the conduct and desires of development professionals. Second, that honesty is required about hypocrisy in working in international development, where differences between values and behaviours become quite apparent. Weaver (2008) has provided a compelling account of hypocrisy in the context of the World Bank, that I have found relevant in reflecting on my own career. Other colleagues have also discussed the use of 'poverty porn'[4] in fundraising and how privileged, pre-dominantly white people maintain their hold on the reins of power, while ostensibly seeking to empower others (Raftree, 2013). To ameliorate the effects of both these lessons the prioritization of the voices of people living in poverty is important because it helps to focus on enhancing development effectiveness in a responsive manner.

Across my career in international development I have experienced development from grassroots to multilateral development agencies over 17 years and have, chronologically, worked for: The Afghan Wireless Communications Company; the University of Cape Town; Volunteer Services Overseas-Regional AIDS Initiative of Southern Africa; the African Medical Research Foundation; the University of the Witwatersrand; the UK Department for International Development (DFID); Oxfam America; and the World Bank. All of these organizations prioritise poverty reduction and social betterment in some form, except for Afghan Wireless, which was a private sector company. The organisations did, however, differ widely in budget, internal culture, location, governance, political outlook, racial and gender

[4] A term that arises from International NGOs use of pictures of children in extremely deprived circumstances, such as famine, to increase fundraising appeals.

composition. People in these organisations navigate these differences in a variety of technocratic, social and political ways.

The most telling difference I have found between the evaluation practices of these organizations is proximity to the field. The first part of my international development career was spent closer to the field, by which I mean in regular contact with people actively looking change the status quo in communities; the second part, in headquarters based in the UK and USA. The experiences in the first part of my career significantly inform my outlook on international development practice and how these intersect with different social forces, such as, class, colonialism, race, gender. Much of the first part, was spent working across Africa, specifically South Africa. Coming from a middle-class upbringing in the UK, the experience of living and working in Afghanistan and South Africa's relatively recent transitions, colliding imperial and ethnic legacies and in South Africa a high degree of both wealth and poverty challenged my political assumptions. I came to better appreciate the importance of struggle and liberation movements, how the deprivations of poverty constrain the lives of people in multiple dimensions and how my birth, race, class and gender often afforded me privileges that by far the majority of people globally could not access. My colleagues who were not from the global north though often motivated by the values of international development, spoke about duty to their family and sometimes their privileged position. Similar to people in other industries the organisational mission was important, but often not the overriding priority in people's life; the values of development sat in tension with individual and group interests.

In implementing monitoring and evaluation systems closer to the field I found that we could, connect to and respond to the reasoning of people we were supposed to serve and work with, an issue Roche (2015) has also commented upon. This had an important grounding effect, helped locate conversations in the contrast between our privilege and the experience of poverty, and helped channel us to often be honest where our own or institutional and personal constraints diverged from our values. Further, what was monitored and evaluated was prescribed or curtailed from headquarters or institutional donors. This could be helpful in that it provided a consistent demand to improve assessment practices, though it could also distract from our own responsiveness to people living in poverty. This conundrum led me to explore how within my career I could reorientate actions to account for marginalized people's voice in the development system by working in a headquarters.

In working in the headquarters of a bilateral donor, international NGO and multilateral organisation in the US and UK I learned to appreciate how the impetus and logic of decision-making responds to proximate structures of power. Whether the organisations were NGOs, government, bilateral and multilateral they struggle to hear and understand the voices of people distant from them who have no formal accountability linkage either by voting or direct representation. In DFID, decision-making was focused on the Secretary of State, in the World Bank towards to board and management, in NGOs towards the HQ based staff and management. While the legitimacy of each of these forms of governance can be debated, the check-and-balance that you get from being close the field is largely absent. With this gap the risk for values to be displaced and hypocrisy increases. Decisions from Oxfam Great Britain, for

example, choose to ignore local governance structures in Haiti and allow abusers of women to continue their careers relatively unfettered (Charity Commission, 2019). In the headquarters while there is some degree of power to open spaces for voice and public reasoning, an agent's existence plays out in a sometimes unhealthy struggle for resources, legitimacy and the focus of attention. For evaluation the incentives in HQ are predominantly positioned to reporting inwards. Experiencing these dynamics led me to continue to explore Sen's capability approach as a theoretical and ethical framework for prioritising voice and seeking to navigate hypocrisy.

2.2 Epistemology

This professional journey enables and constrains the studies that this thesis draws upon. The studies were written as my professional exposure provided rich data and experience to draw upon as a participant observer by reflecting upon the connection between theory and practice. The studies are constrained by my own personal biases and that I was personally involved in developing some of the monitoring and evaluation systems reviewed here. As a result, I perhaps offer simultaneously more generous accounts of the evaluator (myself), and more critical or less informed accounts of other agents. In reviewing these studies for this thesis, I have the benefit of hindsight where claims of success subsequently faded and I acquired reflexivity about my own beliefs, which hopefully helps to partially mitigate some biases. Further, the rigor of the meta-ethnography approach should help wash out some, but not all inherent bias.

The concern with responding to people's voice and interests serves as an important motivation for bringing these studies together and so any particular paradigm that contributes to this end I consider to be valid. In applying the meta-ethnography in this study, I worked through the interpretivist paradigm and sought to stay true to the foundation of the approach and to learn better how to apply interpretation, translation, anticipation and synthesis in a non-aggregative form in a qualitative study. I would look for the knowledge generated here to inform others working in different paradigms. In working towards, comprehending and responding to people's voices and interests, evaluation can apply a wide-variety of paradigms, which could all potentially benefit through a middle-range theory to affect accountability mechanisms.

Though an interpretative approach is taken in this thesis no paradigm is seen as providing the single best way to generate knowledge and understand people's lives. In my journey through development organizations I have found many epistemologies to offer useful insights, as highlighted by Guba and Lincoln (1994). For example, the experimental approaches have offered adaptations to cash transfers and legitimised their wide-spread usage (Lagarde et al., 2007). Whereas participatory approaches have been able to better reveal challenges, for example, with local water governance (Tortajada, 2010). A limitation in working through different paradigms is the recognition that all knowledge produced is partial and will respond to the question being asked, the organization asking the question and the principles informing the research or evaluation. Honest application and interpretation of the different sources of evidence and transparency about the limitations of a paradigm help to better understand the truth of a phenomenon (Murakami, 1989 pp. 125).

2.3 Meta-ethnography and the Interpretive Paradigm

Noblit and Hare (1988: 12) frame meta-ethnography as an approach to synthesise interpretive research, which seeks to explain social and cultural events through peoples' experiences. The paradigm of interpretivism is defined by Noblit and Hare (1988, p. 12) as research that involves thick description (Geertz, 1973), inference, a scepticism of the ability to define enduring laws for social and cultural events and a prioritisation of the experiences of the people being studied. Interpretive approaches interact with different paradigms based on the extent that they are consistent with this definition. For example, an interpretivist paradigm holds that the search for cause-and-effect laws and analysis techniques that focus on aggregation of data have limited explanatory power across social and cultural conditions due to reality being generated through multiple perspectives (Noblit and Hare, 1988). Consequently, taking an interpretivist approach is not just a difference in how data is treated, but a change in perspective about how to understand and synthesise reality so that the context, especially of those whose voices are marginalised, remains salient.

Meta-ethnography was developed when Noblit and Hare (1988) found that their contemporary synthesis techniques lead them to aggregation rather than interpretation of qualitative data. Aggregation of qualitative data commonly occurs, for example, through counting the number of times a phenomenon is mentioned in text strings, as in a framework synthesis (Barnett-Page and Thomas, 2009) or by developing numerical weightings to identify issues of salience in a text (Eisenhardt, 1989). Noblit and Hare (1988, p. 10) work differently, by translating issues between different written texts and in doing so "preserve the uniqueness and holism that characterise qualitative studies." By undertaking translation and interpretation, knowledge is integrated in a manner that maintains the structural aspects and inherent conflicts of the texts. Consequently, meta-ethnography supports the development of middle-range theory by supporting the anticipation of events across contexts, while not aiming at building universal laws.

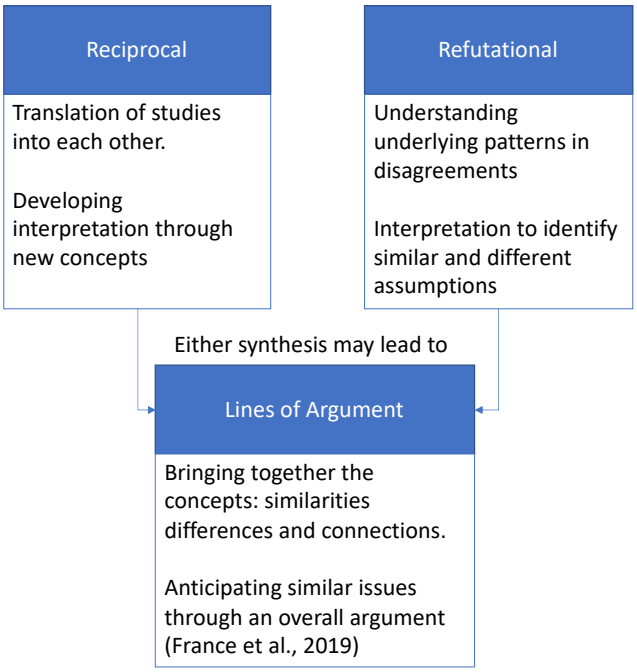
The purpose of highlighting this here is not to claim one paradigm is superior to another nor to attempt a nuanced reconciliation of differing views. Rather it is to recognise that interpretivism provides an alternative way of eliciting insights that potentially differ because of its insistence on maintaining context and nuance in describing events. Meta-ethnography provides a technique to take qualitative insights and synthesise findings in a manner that consistently applies description, explanation and induction to develop important knowledge that can provide an alternative perspective developed through aggregating data.

2.4 Three trajectories of synthesis and seven-phases of meta-ethnography

Applying meta-ethnography within this study builds upon a stable overarching approach that has evolved over thirty years of practice. The approach has been developed to result in three trajectories of synthesis that are achieved through seven-phases that move from identification of an issue through to expressing the synthesis.

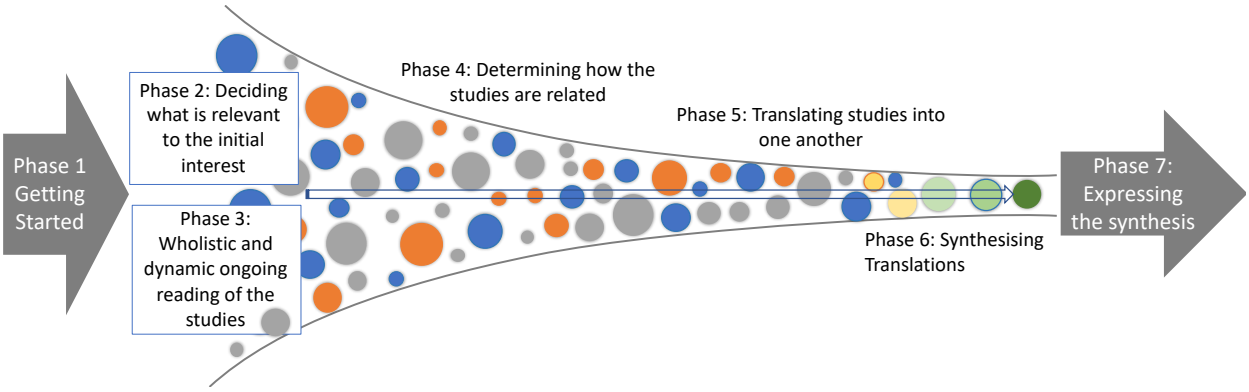
Each of the three trajectories of synthesis for meta-ethnography helps to solve a different issue. First, reciprocal synthesis helps to interpret similar studies when studies are comparable, where "in iterative fashion each study is translated into the terms (metaphors) of the others and vice versa" (Noblit and Hare, 1988, p. 38). Refutational synthesis - the second type of trajectory - attempts to bring together studies that provide competing explanations of an issue. The third type of synthesis develops either a reciprocal or refutational synthesis and is termed a lines-of-argument synthesis. A lines-of-argument synthesis generates middle-range theory by exploring the possibility to say something about the whole and anticipate similar phenomena in different contexts.

Figure 1: Three Trajectories of Synthesis



Conducting these three trajectories of synthesis occurs through a seven-phase process. The seven-phase process is often presented as a list. However, synthesis is more dynamic than sequential, with phases being iterative and overlapping (France et al., 2019; Lee et al., 2015; Noblit and Hare, 1988). Figure 2 provides a perspective on that dynamism of moving from many perspectives to an interpretation that incorporates the whole in reference to the seven-phase process.

Figure 2: The dynamic of meta-ethnography



(Adapted from Noblit and Hare 1988)

Meta-ethnography starts by identifying the research gap (phase 1), which drives the selection of studies (phase 2) – represented by the different coloured bubbles. Studies are then read and constantly reread throughout the synthesis process (phase 3), and juxtaposed with relationships between concepts identified (phase 4). Translations are developed either from the existing texts or new expressions are developed (phase 5). Exploration of similarities and contrasts produces interpretation, which captures relevant qualities from all studies – the final green bubble (phase 6). The synthesis is then expressed in a manner meaningful to the audience (phase 7).

3. Getting Started, the literature review - Phase 1

The connection between evaluation theory and accountability remains under-developed within evaluation literature. Accountability is "one of those Golden Concepts that no one can be against" (Bovens, 2007, p. 448) and has spawned an extensive range of studies. Both the accountability and evaluation literature recognises evaluation as an instrument of accountability and has in limited ways discussed the extent to which evaluation and accountability are entwined (Bemelmans-Videc et al., 2007). Both accountability and evaluation exercise power by seeking to change the actions of agents in a manner that can conflict with their interests. Their entwinement potentially amplifies their effects on agents given that accountability and evaluation reinforce the other. The linkages between evaluation and accountability have yet to be full explored leading to gaps in conceptualisation which have limited the provision of advice on how evaluation affects accountability.

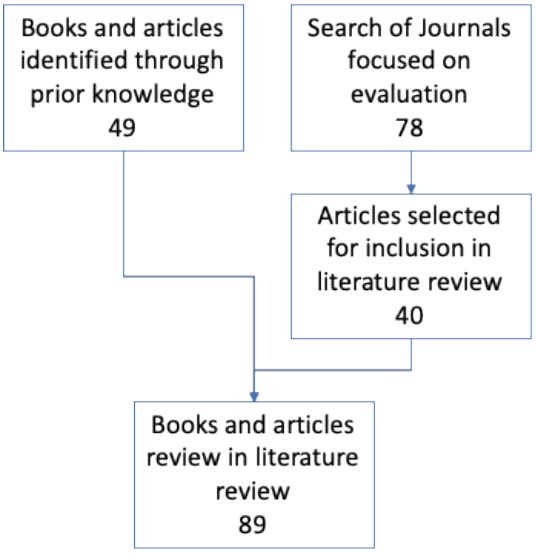
My interest in connecting accountability and evaluation originates in ongoing debates that I have engaged in related to the use and influence of evaluation in organisations in my own professional journey. In re-reading my own work and reflecting on how evaluation could enhance the responsiveness of development practice to the voice and public reasoning of marginalised people I developed a hunch that there was a broader contribution of the articles and chapters to understanding how evaluation affects accountability mechanisms. Based upon this hunch, a literature review was developed, focused on the connection between evaluation and accountability, which then formed the basis for identifying a relevant focus of a meta-ethnography. A vast range of literature on accountability exists that could potentially be drawn upon (Bovens et al., 2014b; Diamond et al., 2005), including studies of accountability deficits (Crack, 2013; Peruzzotti, 2014). The literature on accountability was lightly reviewed as this current study focuses on insights that arise from the perspective of evaluation.

The literature review draws from classic evaluation texts, literature on international development , power, and articles identified in the nine main journals of the evaluation discipline from 2000-2020.[5] The search of the nine main journals applied the term 'accountability' in the title, key word or abstract of the article. This search was conducted to augment the literature already covered in the peer reviewed studies feeding into the meta-ethnography and to develop a further overview of how accountability has been conceptualised by writers of evaluation. The search terms are somewhat blunt, but fulfilled the need to reach saturation on how the discipline of evaluation has understood the interplay with accountability. A total of 78 articles from journals were identified through the search, each article was reviewed for whether they defined and conceptualised accountability, and 40 articles were then selected for inclusion. A further 49 books and articles were identified based upon prior knowledge of the evaluation literature and advice from other experts, outside of the journal search. This resulted in a total of 89 books and articles feeding into the literature review. These were read to understand the treatment of

[5] The journals being: African Evaluation Journal; American Journal of Evaluation; Australasian Journal of Evaluation; Evaluation; New Directions in Evaluation; Evaluation and Program Planning; Canadian Journal of Evaluation; Journal of Multidisciplinary Evaluation; Journal of Development Effectiveness

accountability by evaluation and how the two concepts link to the exercise of power, with emerging themes identified. In exploring the literature, this chapter provides definitions of power, evaluation and accountability, and reviews connections between them.

Figure 3: Literature Review Selection



3.1 Definitions of Power, Evaluation, Accountability and their Connection

Power, evaluation and accountability almost entirely overlap as concurrent phenomenon. Power and accountability inherently entail evaluative exercises, while evaluation entails the exercise of power. Exercising power and processes of evaluation often entail and are reinforced through a connection to accountability, though each can occur with no connection. The concurrency of evaluation, power and accountability is identifiable within their definitions, which are all evaluative, dispositional, and relational. They are evaluative in that they all require value judgments to be made through their processes; dispositional in that they all require a conjunction of conditional statements to occur; and relational as power, evaluation and accountability can only occur in through interaction with other people.

Power

The antonyms contestation and stability best describe debates on the definition of power. A good degree of stability has been achieved in defining power in its general sense, where concern focuses on the extent social actors have power to effect or receive outcomes (Lukes, 2005). In this general view different types of power operate in the conflicts of interests between agents or by consensus (Haugaard, 2002). Five perspectives on the general view of power have been discussed: (i) over others; (ii) within themselves and organizations, (iii) with others, (iv) to undertake an action through a capacity, or (v) under victimhood (Batliwala, 2019). In the general view the capacity for exercising power resides visibly, hidden and invisibly in unquestioned social norms, ideologies, institutional systems, discipline, coercion, tacit knowledge and biases (Gaventa, 2006; Lukes, 2005; Haugaard,2007). This broader view of

power has informed the development of theory and practical courses of action, for example in international development, international relations and public relations (Barnett and Duvall, 2005; Batliwala, 2019; Berger, 2005; Gaventa, 2006; McGee and Pettit, 2019; Reeler et al., 2009). Moving from the general to a more specific view the definition of power becomes contested, for example, between the extent of focus on structure and agency, the possibilities of analysing the active exercise, agenda setting or hidden power, and the applicability of principal-agent conceptions (Hayward and Lukes, 2008; Lukes, 2005; McGee and Pettit, 2019; Oosterom and Scott-Villiers, 2016; Waterman and Meier, 1998).Within these various debates power remains an essentially contested concept because it focuses on different values positions where "reasonable people…, may agree the facts but disagree about where power lies" (Lukes, 2005, p. 63).

Although potentially revealing to analyse the intersection of evaluation and accountability through all types of power (over, within, with, in, under) this task would be complex and require a follow-up synthesis to this one to further test the middle range theory generated here. This follow-up synthesis would draw on studies where different types of power have been reviewed to elucidate important insights on the operation of power (See for example, McGee and Pettit, 2019, pp. 169-212). The studies I authored and co-authored in this meta-ethnography did not consider all types of power in this way. Focusing on one type of power in developing theory and practice is an entry-point to broader discussions of power. As Lukes and Hayward (2008), Fukuyama (2014) and Hobbes (1985) demonstrate, seeking to understand domination alone helps to develop theory. Further as current debates on Black Lives Matter (Clayton, 2018) and COVID-19 (Kissinger, 2020) attest, how domination occurs and how it can work in a beneficial manner, remains a central practical issue for those seeking to reform power and end poverty.

In the context of seeking to construct a middle range theory of how evaluation affects accountability mechanisms this thesis focuses on a specific view of power working from Lukes' conceptualisation. Lukes (2005, p. 110) defines an approach to identify invisible power and answer "how do the powerful secure the compliance of those they dominate?" Accountability and evaluation both work in ways that can bring forth and shape conflicts of interest. As will be detailed later, accountability mechanisms hold the capacity to exercise and constrain power through their defining capability to sanction other interests (Bovens, 2007; Lindberg, 2013). Evaluation, meanwhile, in seeking to inform action needs to internalise and influence the capacity to exercise power, especially to dominate other interests.

A strength of Lukes' approach is the centralisation of the role of agents in shaping the opportunities to exercise power, while acknowledging the role of structures. The agent centred view corresponds also with evaluation and accountability mechanisms. In accountability mechanisms, agents through deliberation make decisions. In evaluation, agents as evaluators assess the results of others' actions. More structural approaches to analysing power according to Lukes (2005), although possibly adding insight to how the structures are shaped would not provide as rich linkage for evaluators and occupants of accountability mechanisms to exercise agency and undertake actions differently.

In accordance with Lukes (2005, p. 37), this thesis defines power in that "A exercises power over B when A affects B in a manner contrary to B's interests." This elegant definition should not be read as over-simplistic, like other one-line definitions its simplicity is an entry-point to working through complexities, in this case of understanding power as domination. Three areas of discussion help to further clarify this definition.

First, this definition should be read as non-binary, recognising that domination will only be over some important interests of an agent and operate in many different ways. Recognising the definition as non-binary emphasises that power will exist in networks of agents, there will be many A's and B's whose relations will be conditioned by social structures that limit or enable different forms of domination and opportunities for resistance (Hayward and Lukes, 2008). Furthermore, within these networks the interests of an agent or many agents are diverse. As Sen (2007, p. 19) highlights he can at the same time hold multiple interests as "an Asian, an Indian Citizen, a Bengali with Bangladeshi ancestry, An American or British resident, an economist, a dabbler in philosophy, an author, a Sanskritist, a strong believer in secularism and democracy." With these multiple interests A could seek to dominate B as the immigrant, but not have power over a range of other interests to induce, encourage or persuade. Consequently, the definition directs us to focus on domination within particular relationships in a broader context, where an adequate counterfactual can be generated to help reveal where interests are in conflict (Lukes, 2005).

Second, applying Lukes' definition engages a three-dimensional view of power that seeks to identify hidden power as a way to better comprehend the capacity to exercise power. A three-dimensional view of power requires focusing on (adapted from Lukes, 2005, p. 29):

- decision-making and control over political agenda, though critiquing a behavioural approach that focuses on visible decisions;
- actual and potential issues that affect the space in which decisions are made, whether through the operation of social forces, institutional practices or individual choices;
- observable (overt and covert), and latent conflict;
- subjective and real interests.

Third, for Lukes (2005) power is a capacity that has the potential to be exercised when a conflict of interests occurs based on coercion, force or manipulation. Agents who hold it can choose to exercise it, as Lukes states (2005, p. 69) "it is a potentiality, not an actuality." As a result, in the study of power analysis is needed of the capacity to exercise power that could be latent and might only be meaningful in a certain context. Without a conflict of interest then the issue at hand could, for example relate to influence, but not power. Finally, power in this definition although involving domination can possibly operate for a broader beneficial purpose.

By applying a three-dimensional view, a researcher seeks to comprehend how invisible sources of power affect the hidden and visible capacity to exercise power. For example, in citing Sen's work on famine Lukes (2005, p. 137) illustrates how following the Bengal famine in India,

women adjusted their preferences and expectations. The women who were considerably more deprived in their health and nutrition status did not consider themselves to be ill and so adapted their preferences. Whereas men in a similar or better situation considered themselves ill and would by implication demand a policy response. Rather than looking at whether stated interests had been achieved Lukes' approach would seek to state a counterfactual situation where absent of cultural, institutional and other conditions that shape women's preferences, they would also consider themselves to be in ill health and so a conflict with their interests would arise. An example of an empirical study applying a thorough three dimensional view is Gaventa's (1982) study of power in the Appalachian Valley that entailed detailed investigation of the relationships of power and their underlying dynamics reaching from the local Unions to the Lord Mayor of London.

Lukes' definition of power is dispositional, evaluative and relational. First, power is dispositional in that it comprises of "a conjunction of conditional or hypothetical statements specifying what would occur under a range of circumstances" (Lukes, 2005, p. 63). For the capacity to exercise power to exist a set of conditions should occur with A in a position to exercise power over B and doing so in a manner contrary to B's interests. Second, power is an evaluative concept. Interests that possibly require the exercise of power are value positions taken by people for moral and political reasons. To comprehend the exercise and distribution of power requires unpacking value positions and understanding how power constrains or enables capabilities. Third, power occurs relationally. Power as domination requires a relationship between A and B. At its core, the study of power identifies how relationships and domination build or crumble through speech, acts, behaviours, interactions, affected by contexts and significant power differences (Hopson and Cram, 2018).

Evaluation

In this research, evaluation is understood to be "a key analytical process in all disciplined and practical endeavours", which focuses upon "determining the merit, worth and value of things" (Scriven, 1991, p. 1). As a key analytical process, evaluation applies to a range of activities including products, projects, programs, personnel, and by extension accountability. Building from this broad perspective, evaluation is an umbrella term incorporating a variety of forms and approaches of assessment, for example, monitoring, impact evaluation, developmental evaluation (Owen, 2007). In conceptualising the implementation of evaluation there are two important components: The first component, the logic of evaluation, provides the overarching meta-theory that is dispositional and evaluative; while the second component, the operational definition, provides guidance for the relational exercise of evaluation and the capacities required in the profession of evaluation.

The process of evaluation proceeds according to a meta-theory, the logic of evaluation, which has wholly or partly been adopted by a wide range of scholars (Fournier, 1995; House and Howe, 1999; Owen, 2007; Patton, 2008; Rossi et al., 2018; Scriven, 1981, 1991, 1995; Shadish et al., 1991) There are four steps in the logic of evaluation, but in practice a fifth step has been added - focused on the use of evaluation - and over the past twenty years this has been further

developed in the literature (King and Alkin, 2019; Kirkhart, 2000; Mark and Henry, 2004; Owen, 2007; Patton, 2008, 2010; Weiss, 1998). The four steps in the logic of evaluation are as follows:

i) Establishing criteria: On what dimensions must the thing being evaluated, the evaluand, do well?
ii) Constructing standards: How well should the evaluand perform?
iii) Measuring performance and comparing standards: What level of performance does the evaluand achieve?
iv) Synthesising and integrating data into a judgement of merit or worth: What is the merit or worth of the evaluand?

These four steps are essentially evaluative and dispositional as each is conditional on the other, all must be performed for the phenomenon of evaluation to be realised and their composition is dependent upon a range of circumstances that provide the context for the evaluation. Fournier (1995, p. 17) has identified the logic within all other types of evaluation, arguing: "It is the basic reasoning that specifies what it means to evaluate something... it specifies the game and the rules of the game that one is playing when conducting an evaluation in any field." By applying these four steps anyone can conduct evaluation to answer questions that they are concerned about, though with variable rigor. From the four steps the potential arises for agents in accountability mechanisms to conduct their own evaluation. Indeed, as will be shown, the definition of accountability overlaps with the logic of evaluation.

The operational definition of evaluation builds from the logic of evaluation by describing the parameters of professional evaluation, in doing so the definition embeds the concept of evaluation use and a range of jargon. The operational definition for international development evaluation originates from the Organisation for Economic Co-operation and Development - Development Assistance Committee (OECD-DAC). The OECD-DAC definition aligns with Patton's (2008, p. 5), three questions that guide evaluation: "What? So what? Now what?" The question of "now what?" highlights the importance of the utilisation of information in the synthesis step of the logic of evaluation:

> "Evaluation is the systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results. [What]
>
> The aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability. [So What]
>
> An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors. [Now What]
>
> Evaluation also refers to the process of determining the worth or significance of an activity, policy or program." (2010, p. 4)

The incorporation of use into an operational definition of evaluation significantly expands the scope of evaluation from the meta-theory. Implementing the logic of evaluation according to the four steps resides within in the control of the agents who choose to do an evaluation. The inclusion of use in the definition - along with jargon such as effectiveness, efficiency and impact - moves evaluation to a highly relational practice requiring capacity as a specialised professional. The assumption here is that the evaluator is independent from the commissioner of the evaluation and needs to interact with other stakeholders in order to improve the likelihood of utilisation of findings. From this definition, the recognition of evaluation as an instrument in accountability arises.

Accountability

The main purpose of accountability is to exercise and constrain power (Bovens, 2007; Lindberg, 2013). Accountability as a concept links historically with the practice and discipline of accounting for value in a monetary sense (Hayne and Salterio, 2014). The concept and practice of accountability has evolved from one where a forum counts expenditures received from the public. Accountability, in expanding its focus on performance and learning, enters a similar terrain to evaluation by seeking to value changes that have occurred in order to inform action. Often the account remains focused on counting, though aspiring to focus on performance and learning (Behn, 2003; Bemelmans-Videc et al., 2007; Bovens, 2007; Honig and Pritchett, 2019).

Bovens *et al*, (2014a) identify a minimal consensus of the definition of an accountability mechanism across a range of literature including linguistic, social psychology, accounting, public administration, political science, international relations and constitutional law. Core to the definition are interactions between those giving the account and those receiving the account, the concept is relational. The minimal consensus on the definition of accountability identified by Bovens *et al* (2014a, p. 6) has four aspects:

i) enabling a relation that links those who owe an account to those who require it;
ii) providing answers to people who can legitimately claim an account;
iii) looking back (ex-post) on what has happened; and
iv) it is consequential, something should happen as a result, with the potential of sanction.

Similar to the logic of evaluation and the definition of power these four aspects are dispositional, relational and evaluative. Each aspect could happen independently; it is only when they arise within a specific disposition, brought together and bonded in processes that the phenomenon of accountability transpires. Accountability operates in a relational manner as a party calls for an account and another set of agents provides answers. Accountability also entails evaluative practices, an issue further unpacked below and in section 3.2.

Beyond the empirical definition, accountability shimmers with a normative framing linked to good governance. Accountability has been positioned academically as "the key to…modern representative democracy" (Schmitter, 2004, p. 18). Within international development the

claim that accountability ensures responsiveness to citizenry burgeons across different accountability frameworks (Bank, 2019; Manroth et al., 2014; UN, 2015). The importance of accountability has arisen because of the increasing fragility of public trust in government activity (Bovens, 2010). As argued by Bovens (2010, p. 954) accountability "in the sense of transparent, responsive, and responsible governance, is meant to ensure public confidence in government and to bridge the gap between citizens and representatives" Consequently, the interest in accountability mechanisms has over 20 years and continues to emphasise the exploration of citizens' voice, perspectives and evidence to engage and influence decision-making and so exercise power in various governance arrangements (Ackerman, 2004; Fox, 2018; Kitschelt et al., 2020; Moncrieffe, 1998; Przeworski et al., 1999).

Differentiation in the application of the concept of accountability occurs between authors who mainly focus on accountability as a virtue and those who have focused on accountability as a mechanism. As a virtue, the analysis of accountability focuses on the behaviour of agents and whether they have undertaken acts that would promote, say, transparency and responsiveness (Bovens, 2010; Bovens et al., 2014a). Viewing accountability as a virtue often means it becomes a synonym for good governance (Schillemans, 2011). In contrast, viewing accountability as a mechanism focuses attention on "an institutional relation or arrangement in which an agent can be held to account by another agent or institution" (Bovens et al., 2014a, p. 9). An agent from an accountability mechanism viewpoint is a person or group who has a specific role to play in a defined process within a specific setting. In this literature, on occasions agents are differentiated from principals. Principals are agents who in a context have a degree of authority. Research on mechanisms provides a view on how the logic of accountability functions, in regards to different agents: who is being held to account, for what, how and by whom (Bovens, 2014).

Fox (2015) in focusing upon social accountability provides a useful elaboration of the accountability as a mechanism perspective relevant to this thesis. A social accountability perspective focuses on citizens and how they are able to attain responsiveness from government, rather than a stated cantered perspective. Fox (2015, p. 347) notes that the principal-agent views of accountability that predominate international development have been stretched and have challenges analysing "non-hierarchical oversight relationships, as in the cases of mutual accountability inherent in partnerships, checks and balances institutions and informal accountability relationships". To counter this shortcoming, he proposes the incorporation of a tactical and strategic distinction to the analysis of attempts to enhance responsiveness to citizens of the state.

The tactical/strategic distinction helps to further unpack attempts to influence reform of accountability mechanisms. A tactical approach to accountability involves a focus on citizen voice, an assumption that information provision alone will inspire collective action and working through local arenas. A strategic approach involves multiple coordinated tactics, work in enabling environments to promulgate collective action, coordination with governmental reforms and iterative processes (Fox, 2015). In reviewing the evidence Fox (2015) notes that a strategic approach to realising responsiveness attains a higher degree of success than deploying

a range of tactics. This distinction is further deployed in this thesis in analysing how evaluation affects accountability as an instrument (tactically) versus a more substantive (strategic) interaction between evaluation and accountability.

There is not necessarily a choice between taking a mechanisms and virtue perspective of accountability, both are applied in this thesis. Taking a mechanism perspective provides a tangible analysis for how accountability works in terms of types, functions, forms and constraints. A focus on accountability mechanisms provides an ability to map the concrete functions of accountability and how they relate to evaluation, in a manner that connects with the everyday practice of evaluation. Taking a social accountability perspective to a mechanism also provides tools to reflect on how citizen voice comes to affect an accountability mechanism. Meanwhile, studying the link between evaluation and the virtue of agents in undertaking accountability provides a lens for analysing the actual and potential behaviours of agents and so identify opportunities for different agents to be affected by evaluation.

3.2 Connecting Accountability with Evaluation

The separate logics of power, evaluation and accountability mechanisms connect based on their overlapping conceptual frames that are dispositional, relational and evaluative. To constrain and exercise power as a form of domination, accountability mechanisms employ evaluative processes, even if they do not receive an evaluation report. The operational definition of evaluation concerns itself with informing action and potentially creating, or supressing a conflict of interests within a given circumstance; and so, evaluators need to engage in a relational manner with an accountability mechanism to support the use of findings.[6]

Within the accountability relationship there is a nesting of the logic of evaluation (criteria, standards, measures and synthesis). In their deliberation, accountability mechanisms conduct all four steps in the logic of evaluation by: questioning the adequacy of information against criteria; and standards; measure performance; and pass judgement to provide a synthesis. The possibility of consequences also overlaps with evaluation. Evaluation is concerned with action; the possible application of consequences is a requirement for a change in actions. The overlapping connection to the logic of evaluation highlights that evaluation is inherent in the process of the accountability relationship.

The connection from the definition of accountability to evaluation and power also arises from requirements for agents to provide information about their conduct. Here the accountability mechanism may require the evaluation to be conducted by an independent professional, to substantiate, nuance or contradict answers provided by an agent. In this circumstance evaluation informs questioning and debate, and is used as an instrument and can put an issue of the agenda and so raise a conflict and exercise power (Porter, 2017).

Bovens (2007), provides a detailed conceptual framework for analysing and assessing the logic of accountability mechanisms, which is unpacked here to further establish how evaluation is a

[6] See Bovens, M. (2007) and Schmitter, P. C. (2004) for discussion of the expansive range of accountability types.

substantive concern within accountability mechanisms. Bovens (2007) provides the base for making the connection between evaluation and accountability because of an accessible frame that is widely recognised. Bovens' approach is augmented by the work of Schillemans (2011) who provides a framework to assess the functioning of horizontal accountability.

There are three elements to Bovens' (2007) framework for analysing accountability mechanisms, that each have a substantive overlap with evaluation: the relational logic of accountability; the broad range of types of accountability; and the effects accountability seeks. Each of these elements will be briefly reviewed below, together they show that in definition and functioning accountability mechanisms have a substantive connection to evaluation.

First, the accountability relationship is formed, when an agent is obliged to explain and justify their conduct (Bovens, 2007, pp. 451, 454). Viewing the logic of accountability as a relation overlaps with evaluation approaches that highlight the relational and interpersonal nature of its conduct (Benjamin, 2008; Chilisa et al., 2015; Chouinard, 2013; Hawkins, 2010; Hopson and Cram, 2018; Kusek and Rist, 2004; Marra, 2018; Mertens, 2009; Patton, 2008; Ryan, 2002). In this literature the emphasis is on how evaluation connects to other agents so that it is conducted in an ethical, legitimate manner that is useful. Writing from the perspective of evaluation Benjamin (2008), for example, specifically explores the relational overlap between evaluation and accountability.

Second, Bovens (2007, p. 454) describes how accountability "comes in many guises" and outlines four questions to consider in mapping accountability mechanisms to identify to whom the account is provided, by who, about what and why. These questions and sub-types provide for the identification of 300 different combinations of accountability mechanism. The list could no doubt be expanded. Recognising the wide-scope and types of accountability emphasises that an evaluation conducted with an organisation will be considered within some sort of account giving mechanism.

Table 1: Types of Accountability

To whom is the account provided?	
Political	Exercise of accountability along a principal-agent chain
Legal	Accountability based on specific legal responsibilities conferred upon authorities
Administrative	Quasi-legal forums conducting independent and external accountability (e.g., auditors and inspectors)
Professional	Accountability to professional peers through an association with codes of conduct
Social	Reporting through public panels composed of external interest groups
Who is the actor giving the account?	
Corporate	An entire organisation provides an account
Hierarchical	Reporting provided along a chain of command, with each reporting to the next layer up

Collective	Each and every member of an organisation account for their behaviour of the whole
Individual	Individuals are held proportionately liable for their personal contribution to the organisation
What is the account to be rendered about?	
Legal	Upholding legal standards
Financial	Meeting financial targets and proprietary standards
Procedural	Following a procedure
Performance	Change contributed to and achieved against benchmarks
Product	The delivery of a tangible output
Why does the actor render an account?	
Vertical	The situation where a forum formally wields power over the actor
Diagonal	Accounting to a forum that sits beside a hierarchy, e.g., to ombudsman or inspectorate
Horizonal	Mutual provision of an account between bodies standing on equal footing

Adapted from Bovens, 2007

In reviewing my own studies using the table above all of them have fed into a type of accountability mechanism. For example, Porter and Goldman (2013), in describing national evaluation systems speaks to a political, hierarchical performance orientated and vertical from of accountability. Whereas Porter (2011) discusses a mainly monitoring system in an NGO that feeds into multiple accountability mechanisms: to donors in a political, hierarchical performance orientated and vertical from of accountability; and internally in professional, corporate, performance and horizonal (see section 4.4 for a full description of the types of accountability mechanisms the studies in the meta-ethnography link to). This typology emphasises that evaluation is very likely to relate to some form of account giving process that has the potential to exercise power, if it is to be useful.

Third, having empirically described the analytical types of an accountability mechanism Bovens (2007: 465 – 466) outlines different effects of an accountability mechanism each of which can be an exercise of power. Accountability can have the effect of democratic control. Here accountability controls and legitimises actions by linking them effectively to the democratic chain of delegation between principals and agents. Success is the degree to which democratic bodies evaluate the conduct of others and enable or constrain agents, so they are responsive to broader constituencies in their actions. Another potential effect of accountability is in providing countervailing power to withstand the ever-present tendency toward power concentration and abuse of powers in the executive. Finally, accountability mechanisms effect learning and improvement. Accountability provides office holders and agencies with feedback and potential inducements or cohesive requirements to increase their effectiveness and efficiency and so stimulate organisations to focus on achieving desirable outcomes.

The effects sought by accountability outlined by Bovens (2007) closely resemble descriptions of the purpose of evaluation that are provided in the academic literature. Learning and improvement overlaps with a purpose of evaluation described by Rossi et al (2018) and Chelimsky (2006). Meanwhile Chelimsky (2006) outlines the accountability purpose of evaluation in a manner closely related to the democratic control and countervailing purposes outlined by Bovens (2007). Chelimsky (2006), for example, emphasises the importance of countervailing powers in her description of the use of evaluation in the US congress. She also highlights the importance of democratic control in principal-agent relations by linking the importance of a well-informed public to undertaking evaluation. These overlaps in the literature highlights that the effects sought by accountability mechanisms do align to the purposes of evaluation.

The framework presented by Bovens, further reinforces that the logic of accountability mechanisms and evaluation are entwinned at conceptual level and that both are relate to the capacity to exercise power. The relational underpinnings sets-up a dynamic through which power can be exercised and augmented by the use evaluation products and processes. Given the range of accountability types, it is foreseeable that almost all evaluations will be considered in an accountability mechanism, though use can take place in other forums. The effects sought by accountability mechanisms and in their capacity to exercise power overlaps with the purposes of evaluation. The consistent overlap provides a basis for further exploring how evaluation views and affects accountability.

Beyond this section the scope of arguments and reflections on accountability are mainly concerned with how evaluation practice views and connects to accountability, reflecting the research gap being investigated. Though this thesis focuses mainly on evaluation, the grounding provided here demonstrates that there are overlapping concerns on with evaluation from the perspective of accountability that provide perspectives and language that inform how evaluation connects to accountability, though further research would be beneficial to expand the scope of the following discussion.

3.3 Accountability Overlooked by Evaluation Theory

The substantive conceptual overlap in the logics of accountability and evaluation is not explored in the literature on evaluation. In reviewing the literature of evaluation, two main perspectives and one partial perspective emerge on how to treat accountability. Evaluation literature either makes no explicit connection or makes a connection where evaluation works as an instrument for use within accountability. A limited part of the evaluation literature develops an incomplete view of how evaluation affects the reasoning of accountability mechanisms. Looking across this literature some common concepts emerge around how evaluation and accountability connect, that can be further developed.

Rogers (2005: 2), in *The Encyclopaedia of Evaluation*, issues a challenge to improve the way in which evaluation conceptualises its link to accountability, arguing that "the ways in which evaluation is used for accountability are frequently so poorly conceived and executed that they are likely to be dysfunctional for programs and organisations." Answering this challenge

requires us to step back and rethink how evaluation and accountability are and could be connected, based on overlapping reasoning.

Some evaluation literature has an absence of an even a minimally developed connection with accountability. For example, Pawson and Tilley (1997), Shadish, Cook and Leviton (1991) and Mertens (2009) do not make connections even where there are potential linkages. In all of these texts a thicker conceptualisation of accountability mechanisms could help to further refine their proposed evaluation approach.

Pawson and Tilley (1997), do not mention the term accountability at all in their text *Realistic Evaluation*. In proposing the realist evaluation approach, they highlight that they seek to accumulate knowledge over successive trials of a programme, but not how use happens. Pawson and Tilley (1997) do highlight that evaluation can be geared to decision makers' objectives, but this argument is not elaborated upon. An understanding of accountability would help to think through how evaluation needs to connect to decision-makers in order to support knowledge accumulation.

The book *Foundations of Program Evaluation: Theories of Practice* by Shadish *et al* (1991), is a seminal work in evaluation theory development. The book defines the components of a good theory of program evaluation based upon a review of seven leading North American-based evaluation theorists. The authors identify five components of good theory for program evaluation, namely: knowledge use; social programming; valuing; knowledge construction; and practice. Within Shadish *et al* (1991) accountability is referenced in their review of different theories of evaluation. There is a disconnect, however, in their final synthesis, with no direct mention made of the connection between evaluation and accountability in the five components. For example, none of the five components of good theory for program evaluation identify accountability as an area of importance. In similar vein to Pawson and Tilley, accountability is overlooked.

Mertens (2009, p. 3), outlines an approach to evaluation "in response to individuals who have been pushed to the societal margins throughout history and who are finding a means to bring their voices into the world of research." In this text, although advice is provided on how to work and interact with decision makers, the approach to working with accountability mechanisms is not outlined in a substantive form. The lack of reflection on this interaction represents a gap for Mertens. The axiology of the transformative paradigm is defined in terms of human rights and social justice; both strongly rely upon and seek to constrain the exercise of power. As the exercise of power is the main concern of the transformative evaluation approach, then it would be helpful to elaborate the connection to the accountability mechanism.

3.4 Evaluation Literature that Makes a Partial Connection with Accountability

Some literature on evaluation has sought to clarify connections with accountability and defines accountability as a key purpose of evaluation. Fulfilling the accountability purpose of evaluations occurs when agents apply evaluation as an instrument for an accountability

mechanism. Beyond the academic literature, the view of evaluation as an instrument for accountability is reflected in international development evaluation practice. This section first provides an overview of some of the main way's evaluation has been conceptualised as an instrument in accountability. The section secondly discusses evaluation literature that focuses on the tensions between accountability and another purpose of evaluation, learning. Third, literature is reviewed that grapples with providing evidence to inform international development practice, within systems of power that constrain responsiveness to poor people's voice, often through evaluation. The final part of this section describes how some evaluators have built frames that suggest that evaluation affects accountability in a substantive manner beyond its application as an instrument, an argument only partially developed in these texts.

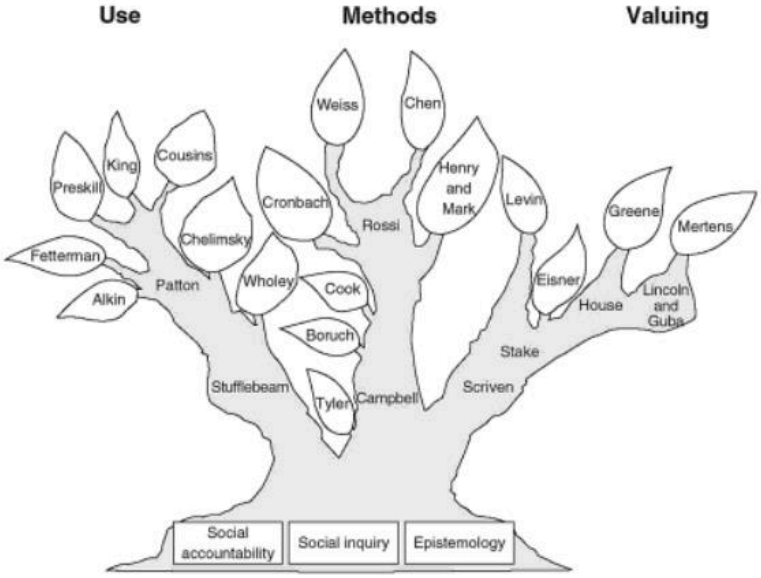
Evaluation as an Instrument for information, transparency and summative judgement in Accountability

Three texts - Christie and Alkin (2013), OECD (2010) and Rossi, Lipsey and Henry (2018) - illustrate different strands of argument around evaluation working as an instrument in accountability mechanisms. Christie and Alkin (2013) emphasise the delivery of information, the OECD (2010) adds transparency, while Rossi *et al* (2018) focus on evaluation providing summative judgement.

Christie and Alkin (2013, p. 15) place Social accountability[7] as one of the three roots in their evaluation theory tree, the two others being systematic social inquiry and epistemology. As represented in Figure 2 below with the corresponding authors.

Figure 4: Evaluation Theory Tree

[7] In an earlier version Alkin and Christie Alkin, M. C., and Christie, C. (2004). An Evaluation Theory Tree. In M. C. Alkin (Ed.), *Evaluation Roots: Tracing Theoristsí Views and Influences* (pp. 12-65). Sage. simply referred to 'accountability'. The later version of accountability seems to refer more to what Bovens Bovens, M. (2007). Analysing and Assessing Accountability: A Conceptual Framework1. *European Law Journal*, *13*(4), 447-468. https://doi.org/10.1111/j.1468-0386.2007.00378.x would call public accountability, in that the public has a stake in the account, which differentiates social accountability as a process in which the public engages.



Source: Christie and Alkin. 2013

Christie and Alkin (2013, p. 15) argue that the social accountability root situates and legitimates "evaluation as a fundamental process for generating systematic information for decision making" and that "evaluation simply provides the information for "being answerable."" From an evaluation theory perspective, they are arguing that accountability is a key area for evaluation use, but only in so far as evaluation is an instrument to feed information to accountability mechanisms. There is limited elaboration beyond these main points that make an initial connection between evaluation and accountability. Christie and Alkin (2013), for example, do not make any arguments about the substantive role of evaluation in shaping the functioning of an accountability mechanism.

From a development evaluation perspective, the OECD (2010, p.7) in the *Summary of Key Norms and Standards* highlights that in addition to learning, the main purpose of evaluation is "to provide a basis for accountability, including the provision of information to the public". This definition emphasises transparency of information in order to provide answerability to the public.

The partnership aspect of evaluation cited within the OECD norms and standards suggests a more substantive interaction between evaluation and accountability. The discussion of a partnership approach proposes that evaluation can help to build mutual accountability and undertake capacity development which is supportive of an environment of accountability (OECD, 2010: 21). However, this view of evaluation and its connection to accountability, which goes beyond the provision of information, is not developed further.

Rossi *et al* (2018), meanwhile, state that the four main purposes of evaluation are improvement, accountability, knowledge generation and hidden agendas. They describe evaluation as an instrument for providing summative judgement to contribute to significant

decisions about the continuation of a program. In focusing on the summative decisions, they appear to be defining accountability as synonymous with good governance, a virtue that is exercised in the behaviour of an actor, rather than as a mechanism that can be shaped. The use of evaluation for summative purposes is then a demonstration of virtuous behaviour: The program has been held to account.

Rossi *et al* (2018) hint at a more substantive role for evaluation in accountability in reference to stakeholder relationships. Rossi *et al* (2018, p. 14) recognise the potential of an empowerment evaluation for "not only producing informative and useful findings but also at enhancing political influence of the participants." The statement flags a potential entry-point for evaluation to affect the design and actions of an accountability mechanism, rather than merely informing it. The potential highlighted for evaluation to affect accountability mechanisms is not developed further in the text, perhaps because the frame for connecting evaluation and accountability has already been limited to summative judgements.

These three texts establish that evaluation does play a role as an instrument within accountability and hint at a broader case for how evaluation affects accountability. The texts consistently highlight that good evaluation should be defined as an instrument that supports accountability mechanisms to make consequential decisions. Evaluation as an instrument is different to evaluation playing a substantive role in accountability mechanisms, however. A substantive role recognises that evaluation can affect the way an accountability mechanism approaches partnership, learning and social justice. Each of these texts hint at the substantive role, but do not develop the idea.

Reconciling accountability and learning purposes of evaluation – imposed reality, follow-up and differing values

Debate about the tensions between the learning and accountability purposes of evaluation endures (Picciotto, 2018). The evaluation literature reviewed below has moved some way in providing advice and ways to think about bridging the accountability and learning purposes of evaluation. The literature surveyed here seeks to reconcile the tension between the accountability and learning purposes found in practice, by adjusting the role that evaluation plays. Discussion of the substantive connection between evaluation and accountability also remains under-developed in this literature, however.

The tensions between the accountability and learning purposes have been noted to be especially acute in settings where evaluation links those who allocate funding and those who receive funding (Guijt, 2010). Kachur *et al* (2016) discuss how non-governmental organisations in international development often perceive the only function of conducting evaluation as serving the imposed reality of donor accountability. Kachur *et al* (2016) argue that in order not to be overwhelmed by external accountability requirements, it is important to develop an internally generated organisational view of the information required for learning and change. Kachur *et al* (2016) also argue that the development of internal information requires capacity for evaluation and the implementation of participatory techniques. Arguments along the same line are made by Lennie and Tacchi (2014), Guijt (2010) and Porter and de Wet (2009) in

international development, and by US based authors about education accountability (Chouinard, 2013; Greene, 1999; Patton, 2010; Ryan, 2002; Ryan, 2004, 2005).

Picciotto (2002, 2003, 2018) and Feinstein (2012) argue that accountability and learning go hand-in-hand, with an evaluation process addressing each. Evaluations in providing feedback support the learning purpose, whereas follow-up on findings from evaluations requires accountability. The articles by Picciotto and Feinstein also highlight the importance of accountability mechanisms for incentivising learning. These arguments recognise that accountability mechanisms produce demands through their criteria to influence evaluation and learning processes. For example, if an accountability mechanism demands evidence on whether an organisation learns about performance, an evaluation may well be commissioned to dig deeper into this issue. In this view, evaluation provides an instrument for both learning and accountability processes, with each process mutually reinforcing the other. Picciotto (2003) tantalisingly takes this argument further by highlighting that when evaluators transparently speak truth to power, decision-makers can learn and potentially shift their criteria. Beyond this statement, how the evaluator affects demands and incentives of accountability mechanisms remains undeveloped in these articles.

Tensions do emerge with learning when evaluation works as an instrument for accountability mechanisms. Some of the main issues that give rise to the tension between the accountability and learning purposes of evaluation are the challenges of managing external funding arrangements, a lack of capacity to demand information in internal organisational processes, and insufficient participation in the evaluation process.

Grappling with Power and Providing Evidence that Informs Practice

Practitioners and academics have grappled with how to provide evidence to inform development practice, conscious that systems of power constrain the responsiveness to poor people's voices. This literature emerges from efforts to identify pathways for evidence informed policy and practice that are grounded in poor people's experiences.

The literature on power and evidence based-practice focuses mainly on *tactical* adaptations to practice and is often limited in identifying *strategic* opportunities for interactions, applying Fox's (2015) terminology outlined in Section 3.1. Practitioners linked to international NGOs identify how they can operate with evidence in a politically informed manner in their own organizations with limited reflection on how they shift donor environments (Eyben et al., 2015; Hayman et al., 2016). Evaluators provide important methodological approaches to expanding voice and authenticity with limited consideration of who they report to in accountability mechanisms (Bell and Aggleton, 2016; Cornwall and Aghajanian, 2017; Johnson and Rasulova, 2017). Some evaluation scholars recognise constraints in larger political structures though their advice remains largely focused on tactics undertaken in the process of evaluation (Hedrick, 1988; Weiss, 1993). Some more recent literature, however, has started to think about evaluation practice in broader political ecologies. For example literature focused on institutionalizing demand for evidence provides theory for how to interact with policy

processes, though not addressing accountability mechanisms specifically (Goldman and Pabari, 2020; Hopson and Cram, 2018).

Experienced practitioners have engaged in collective reflections that brought together thinking around evidence, research, and knowledge in international development related mainly to northern INGO practice (Eyben et al., 2015; Hayman et al., 2016). Both the work of Eyben et al (2015) and Hayman et al (2016) arise out of disquiet about how the term 'rigour' and 'results' had been distilled with accompanying concerns about negative power dynamics that excluded perspectives that did not conform to causal quantitative impact evaluation approaches. Based on the review of case studies provided by practitioners both texts concluded with quite similar synthesis reflections. The summary reflections of Hayman et al(2016) and Guijt (2015) emphasise employing tactics, such as, developing experiential knowledge focusing on people and culture with NGOs playing a brokering role to surface this experience, creating space for learning, negotiating knowledge within political constraints, asserting organizational values and working with the positive aspects. Both texts also start to envisage a more strategic space for political action that tries to shift the discourse in decision-making structures within NGOs, yet these ideas are not developed into a broader theory of engagement with structures of power.

Another set of academics who practice evaluation have sought to improve the responsiveness of development through the implementation of methods that employ and interpretivist approach. Bell and Aggleton (2016), Cornwall and Aghajanian (2017) and Johson and Rasulova (2017) all provide important methodological descriptions that show how qualitative work grounded in an interpretive tradition can add depth to comprehending international development work. The authors show how focusing on authenticity or participatory approaches enable answers to how, why, when and where questions that move beyond the attempts to answer questions in a more aggregative quantitative approach. Within these approaches they illustrate how to enable voice to emerge and operate reflexively on their own positions of power. The papers, however, do not move much beyond discussing evaluation implementation and so provide a tactical approach and a set of revised instruments rather than a strategic approach to shifting the functioning of accountability mechanisms.

Academics and practitioners have identified how evaluation interacts with power as a form of domination and offer tactical advice to evaluators in adapting process or to the context, but do not propose theory in helping accountability mechanisms to adapt. Hedrick (1988) described how power through politics shapes the scope of evaluation, the findings, the use and dissemination of results and how evaluation helps to provide a forum to bring disagreement forth in a tangible manner. The advice of Hedrick on resolving these issues focuses on process issues around communication and the formation of advisory groups. Weiss (1993) discusses how programme managers operate in a different logic and perceive evaluators as a threat to programme survival rather than valuing systematic evidence. Further Weiss (1993) discusses how evaluation within the higher decision-making space clarifies trade-offs, with the evaluator being successful when they closely align to decision-makers values. Weiss does not, however, take this insight further to discuss how evaluators could work with decision-makers to shift their assumptions. Similarly, Chelimsky (2006) in describing her work in Congressional oversight

identifies how evaluation can respond and adapt to political power, but does not offer advice for shaping changes in accountability mechanisms.

Recent publications (Goldman and Pabari, 2020; Hopson and Cram, 2018) go beyond the identification of tactics and offer to differing degrees strategic frameworks for working towards evidence addressing societal issues with a concern for marginalized voices and responsiveness. In doing so they provide entry-points to substantively affecting accountability mechanisms. Goldman and Pabari (2020), provide a realist styled analytical framework that seeks to achieve behaviour change in the institutional policy structures based on case studies of evidence-based practice in Africa, by African-based researchers who have worked with governments. What makes the framework strategic is the emphasis of working to affect political policy making through active facilitation and knowledge brokering, developing formal structures and systems to maintain ownership, and developing the capacity of managers and decision-makers (Goldman and Pabari, 2020, pp. 235-239). In providing the framework they highlight the importance of developing relationships and analysis of opportunities as power structures shift. These forms of consideration to not arise in the literature reviewed above and can inform the narrower focus of this study on how evaluation affects accountability mechanisms.

Hopson and Cram (2018, p. 9) argue that evaluation in complex ecologies where systems of power marginalize ecological and indigenous perspectives needs to attend to 3Rs: relationships, relevance and responsibilities. This argument draws upon cases that provide lessons from ecological, indigenous, ethnographic, anthropological and policy perspectives. The book argues that in forming relationships evaluators need to identify and relate to key interests, points of interaction, variables and stakeholders amid dynamic and complex issues in an honest manner that is based on building trust in evaluation. Responsibilities involve acting with propriety, doing what is proper, fair, right, just in evaluation against standards. Relevance regards being accurate and meaningful technically, culturally and contextually. These 3Rs provide a framework that helps knit together strategic and tactical concerns; working to the 3Rs helps to coordinate multiple tactics, both with citizens and political decision-makers. Tuhiwai Smith (2018) working from a Māori provides a strategic perspective in working with the 3Rs in discussing how the interface between evaluation and accountability mechanisms can maintain colonial knowledge systems. She highlights a focus on working through the 3Rs helps to co-create evaluations strategically in a culturally relevant manner that helps to "*be in relation*" (Tuhiwai Smith, 2018, p. 63) with other entitles.

The literature reviewed connects power with evidence informed practice, recognises the challenge of addressing power, identifies approaches to open spaces for voice and moves beyond a view of evaluation as an instrument in decision-making. Yet the literature often provides a series of tactics to stretch practice through changes in practice and improved efforts to connect with marginalized people's voices. In the main a coherent strategic approach for influencing power does not emerge that coordinates actions at different levels (Fox, 2015). The exception is more recent work by Goldman and Pabari (2020) and Hopson and Cram's (2018) who knit together tactical and strategic concerns in undertaking an evaluation. These two approaches highlight that evaluators can work in strategic and tactical engagements that

coordinating action with decision-makers and so provide useful guidance on substantively engaging with accountability mechanisms.

Exploring how Evaluation affects the operation of Accountability

Authors Fetterman and Wandersman (2005), MacKay (2007) and Behn (2001; 2003) have identified the potential for evaluation to shape an accountability mechanism. This literature highlights how evaluation processes can affect accountability mechanisms through shifting the roles agents play, incentives and the criteria and standards applied to success. Yet these insights are not developed into a middle-range theory.

Fetterman and Wandersman (2005) in describing Empowerment Evaluation outline ten characteristic principles[8], with accountability being their final principle. They define empowerment evaluation as an approach that seeks to improve programme success through "tools for assessing the planning, implementation, and self-evaluation of their program, and… mainstreaming evaluation as part of the planning and management of the program /organisation."[9] The authors argue that empowerment evaluation mainly interacts with internal accountability mechanisms that are built into an organisation's structures. The authors contrast internal accountability with external accountability to a funder, which only lasts "as long as the external agency is present to exert its force" (Fetterman and Wandersman, 2005, p. 50). Fetterman and Wandersman (2005) argue that accountability shifts from supervisor driven to shared agreements, resulting in self-evaluation processes. Stakeholders included in the evaluation process then hold each other accountable to these shared agreements, which may include program participants. The shift to shared agreements happens "because the work is in alignment with individual and group interests" (Fetterman and Wandersman, 2005, p. 50).

Fetterman and Wandersman's (2005) work highlights that processes of participation, collaboration and developing evaluation capacity support a change in the accountability mechanism as well as the strength of an evaluation process. The insight that evaluation shifts the behaviour of stakeholders engaged in accountability mechanisms is not developed further. Although the authors provide examples of accountability shifts within their own practice, they do not elaborate how their approach to shifting accountability could be utilised in other evaluation approaches that are not empowerment focused. Further, many of the examples of empowerment evaluation practice are in reasonably contained settings, rather than across large sprawling bureaucracies. As a result, though there are lessons to draw, there are also limits to the generalisability of their experience.

---

[8] The ten principles in order are (1) improvement; (2) community ownership; (3) inclusion; (4) democratic participation; (5) social justice; (6) community knowledge; (7) evidence-based strategies; (8) capacity building; (9) organisational learning; and (10) accountability.

[9] The definition provided by Fetterman and Wandersman Fetterman, D. M., and Wandersman, A. (2005). *Empowerment evaluation principles in practice*. Guilford. that focuses on capacity and assets contrasts to more feminist definitions of empowerment. See, for example, Cornwall's Cornwall, A. (2016). Women's empowerment: What works? *Journal of International Development*, *28*(3), 342-359. discussion of the evolving definition of empowerment.

Mackay's (2007) writing is based on his experience in building country evaluation systems in development settings with the World Bank, and offers insights into working with accountability mechanisms at a broader level. Mackay (2007) draws upon a rich institutional history of the World Bank in exploring the demand for evaluation. Picciotto (1995, p. 23), based on experience from the Independent Evaluation Group of the World Bank, laid down the challenge: "Say's Law does not fully apply: good evaluation does not always create its own demand. Demand must be nurtured in parallel." Though the challenge is not fully answered in Picciotto's article, other writers drawing on experience from the World Bank offer further insights. Shah (1998), for example, emphasises the role of an active citizenry in shaping the demand for evaluation; while Picciotto and Wisener (1998) and Wisener (2011) discuss the role of incentives in shaping demand. Mackay (2007) links to these arguments and further builds out the frame for evaluation as an instrument in accountability mechanisms.

Mackay goes beyond a focus on an instrumental view in two important lines of argument. First, Mackay (2007, p. 73) identifies four different purposes of monitoring and evaluation focused on policy processes that all exercise power and are implemented through accountability mechanisms, namely:

i) to support budget decision making or national planning;
ii) to help ministries in policy formulation, policy analysis, and program development;
iii) to support ministries and agencies in managing their activities;
iv) or to strengthen accountability relationships.

Other authors, such as House and Howe (1999), do discuss the role of power in evaluation, but are more concerned with how to mitigate its role in the process of conducting an evaluation. By contrast, Mackay's four purposes of evaluation articulate the role of evaluation as exercising and constraining power in governmental decision-making.

The second line of argument regards the incentives and the demand-side of evaluation (Mackay, 2007). Working to encourage evaluation demand entails the definition of incentives – termed carrots, sticks and sermons - to encourage people to commission and undertake good quality evaluations (See for example,Bemelmans-Videc et al., 2003; Funnell and Rogers, 2011). In meeting demands evaluation is argued to be more likely to be used as an instrument in decision-making, an argument also highlighted by other authors (Boyle and Lemarie, 1999; Jasanoff, 2004; King and Alkin, 2019; Packard Foundation, 2010a; Wiesner, 2011).

The main contribution of this work is to highlight that people's demand for evaluation does not originate in a vacuum. By taking this into account, accountability mechanisms can be better understood and incentives developed to use and take evaluation seriously (Mackay, 2007). The shaping of incentives to meet demand implies that the agenda of accountability mechanisms can be affected by evaluation, an issue also highlighted elsewhere in the evaluation literature (Boyle and Lemarie, 1999; Mayne et al., 1999). Mackay, however, does not fully develop a frame for how evaluators can work to affect the broader incentives deployed by accountability

mechanisms. The concept of demand for evaluation has the potential to be deployed and understood in an expanded manner, an issue that is explored in my own work (Porter, 2011; S Porter, 2013; Porter, 2017; Porter and de Wet, 2009; Stephen Porter and Ian Goldman, 2013; Porter and Hawkins, 2019) and synthesised within the meta-ethnography of this study.

Behn (2001; 2003), a writer focusing on accountability from a public administration perspective, offers another partial perspective on the substantive role that evaluation can have in an accountability mechanism. In making this argument he highlights the role of evaluation as an instrument in accountability, arguing that to hold a public agency accountable they first have to evaluate its work (Behn, 2014). Behn (2001) further highlights how criteria and standards are set *within* an accountability mechanism, rather than a mechanism simply using evaluation. Behn (2001, p. 10) argues that to "specify the level of performance we expect of a public agency, we need…a clear benchmark of performance. ...To establish our expectations for what a public agency will accomplish…we citizens need to specify the results we want it to produce." The establishing of benchmarks and expectations for performance within an accountability mechanism is an evaluative act, of setting criteria and standards that establish how the accountability mechanism functions. This signals the potential of a more expansive role for evaluation in accountability - drawing upon the logic of evaluation to shape how an accountability mechanism operates, as well as being an instrument.

Behn is not alone in making an argument that starts to connect evaluation and accountability in a substantive manner. This view overlaps with Bovens (2007) in emphasising that evaluation is not only an input, but an inherent part of the process of accountability. Lindberg (2013) also argues the importance of standards and measurable expectations in the exercise of accountability. Although these statements move beyond much of the evaluation literature by identifying the overlap between the logic of evaluation and accountability mechanisms, these arguments are not fully developed. They hint at a wider role, but do not present an empirically founded framework for how to get an accountability mechanism to undertake an evaluative role.

Beyond these three authors, others have pulled at the thread that connects evaluation and accountability. Schoenefeld and Jordan (2019) and Stame (2003) have discussed the importance of considering the political opportunity dimensions in how accountability and evaluation connect. Rutman (1990) highlights how management and political accountability mechanisms differ in the extent to which they are clear about the information they require. Van Der Meer and Edelenbos (2006) describe how multi-actor policy processes require evaluation to interact with more networked forms of accountability, requiring the evaluator to foster greater co-operation between actors. Zapico-Goñi (2007) has argued that the criteria applied by both evaluation and accountability mechanisms should support adaptation to help manage programs in uncertain contexts. Similar to the literature discussed previously, these arguments provide less information on *how* the connection happens than what could be changed.

3.5 Sketching Areas of Connection between Evaluation and Accountability

The connection and interaction between evaluation and the implementation of accountability mechanisms is yet to be fully developed. Pathways for making a fuller connection do emerge in the literature review undertaken here (for example, Behn, 2001; Fetterman and Wandersman, 2005; Mackay, 2007; Hopson and Cram, 2018; Goldman and Pabari, 2020). There are three issues that start to emerge from the literature review to further develop.

First, power, evaluation and accountability almost entirely overlap as concurrent phenomenon. Accountability applies a logic of evaluation within its processes, while evaluation requires a connection to accountability if its aspirations of changing action are to be fully met. Both accountability and evaluation entail the exercise of power over others. The definitions of all three are evaluative, dispositional, and relational. These connections can be leveraged through alignment in evaluative criteria, in working with decision-makers so that they can perceive the unity of the three in their actions and ensuring that the context remains in focus.

Second, tensions exist that can be resolved. The tension between learning and accountability can be addressed through developing the demand for evaluation, aligning incentives, developing capability of internal decision-making processes and further participation in the evaluation process. Operating within structures of power that reduce voice can be addressed through tactics that develop new methodologies, experiential knowledge, adapt the people and culture in organisations, or surfacing this experience of disjuncture between stated values and practice.

Third, evaluation practice provides an instrument of accountability and strategic approaches to affect accountability mechanisms. Evaluation provides evidence that feeds into an accountability mechanism, helping to form an account of what has happened. Evaluation processes are highly likely to work through a type of accountability mechanism due to their range in contemporary organisations. In the literature there are hints of the potential to influence the values of accountability mechanisms through evaluation, with the literature emphasising participation, tactical approaches to shift practice, understanding the demand for evaluation in a broader political environment, and engaging on the definition of performance.

4. Deciding what is relevant - Phase 2

Evaluation and accountability relate but how evaluation practice affects accountability mechanisms is not fully understood. Thus, evaluators and those managing evaluation systems would benefit from a middle-range theory focused on this area, so that evaluation systems can be better aligned to accountability mechanisms. Evaluators through a middle range theory would have an increased range of conceptual frames to understand how objectives of social betterment (Mark et al., 1999), oversight (Chelimsky, 2006), transformation (Mertens, 2009) or performance (Rossi et al., 2018) are constrained by accountability mechanisms. For agents managing an accountability mechanism, a middle-range theory would help them to understand how to interact with evaluation. In the international development field a middle-range theory could identify especially pertinent issues, given concerns about how power is exercised through accountability between those that provide assistance and those that receive it (Crush, 1995).

The five studies selected in Section 4.4 which form the basis of the PhD originate from undertaking evaluation in the field of international development. At a time when institutions are being tested by the COVID 19 pandemic and misinformation, it is important to have accountability mechanisms that are not only informed by evaluation, but also conduct good processes of evaluative reasoning. In this context understanding how accountability mechanisms and evaluation can interact to achieve shared goals through the use of reliable information usefully contributes to pressing issues of the day. Before presenting the studies chosen for meta-ethnographic review, I briefly state the overarching question and objectives of the PhD.

4.1 Question and Objective

The main question of this study is: How does evaluation affect accountability?

The objective of the PhD research is to: develop a middle-range theory of how evaluation affects accountability mechanisms based on a synthesis of existing research. To achieve this objective, this thesis builds upon the literature review by conducting a meta-ethnography of narrative accounts that I have published in peer-reviewed journals that review the interaction between evaluation and accountability in different settings through empirical studies.

4.2 Selection of Studies for In-Depth Meta-Ethnographic Review and their Coherence

In re-reading my own work and its bibliography generated over 10 years of study of evaluation, I developed a hunch that a broader contribution could be drawn from the articles and chapters towards understanding how evaluation affects accountability mechanisms, a gap that was confirmed in the literature review. To fill this gap, I returned to my own studies where the interaction between accountability mechanisms and evaluation is consistently explored, to see if there were any new insights that might emerge by considering the portfolio as a whole through meta-ethnography.

I employed a purposeful search strategy to identify studies out of my nine peer reviewed publications. The use of the purposeful search strategy is justified, since the basis of the meta-ethnographic approach is founded upon and continues to utilise purposive samples. Noblit and Hare (1988) developed meta-ethnography using a purposive review of school integration studies in the United States. The tradition of purposive selection has been continued with recent meta-ethnographic studies and reviews that provide advice on working from a purposive selection of studies (Campbell et al., 2011; Campbell et al., 2003; Pilkington, 2018), which were utilised in designing the methodology for this study. France et al (2019), who provide reporting guidance, recognise purposeful search strategies as a legitimate means for meta-ethnography, based on an earlier methodological systematic review. Examples of meta-ethnographic reporting were also referenced to develop this study and revisited on an ongoing basis (Atkins et al., 2008; Britten et al., 2002; Campbell et al., 2011; Campbell et al., 2003; France et al., 2014; Malpass et al., 2009).

The use of my own studies is justified because of the limitations of current empirical research in the discipline of evaluation, where theory is produced based on reflections of practice, often with limited transparency (Blake and Ottoson, 2009; Miller, 2010; Ottoson, 2009; Patton, 2008; Shadish et al., 1991; Weiss, 1998). Improving evaluation entails asking how and why the evaluation process unfolded as it did, which is often very qualitative in nature. Evaluation theorists guided by underlying assumptions of their paradigm, reflect upon their practices and develop evaluation approaches (Pawson and Tilley, 1997). Authors who have developed evaluation approaches, provide an array of information to help implement an evaluation, but little on how the approach and advice was developed. For example, guiding texts on developmental evaluation (Patton, 2010), transformative research and evaluation (Mertens, 2009) and program evaluation (Chen, 2005)[10] tells you about the why, how and what of the practice recommended, but little on how practice and paradigm were synthesised to develop advice. Further codification of how to synthesise knowledge on evaluation practice would support discussion on the quality of evaluation theory and the identification of overlaps.

Selecting studies well known to the author fits well with the intent of meta-ethnography. Ethnography focuses upon thick description, participant-observation, understanding an issue from the position of people's lived experience (the emic) and through this providing interpretation. Synthesis requires interpretation and understanding of contexts in which phenomena emerge. Moving from thick description to synthesis and summary can fail,

[10] Developmental evaluation Patton, M. Q. (2010). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. The Guilford Press. is an approach to evaluation that applies complexity concepts to enhance innovation and the use of evaluation. Transformative research and evaluation Mertens, D. M. (2009). *Transformative research and evaluation*. The Guilford Press. prioritises the furtherance of social justice and human rights through community involvement in the methodologies and practice of evaluation. Program evaluation is focused upon specific interventions that seek to achieve an overall purpose e.g. HIV stigma reduction or development of social work competencies. Chen's Chen, H.-t. (2005). *Practical Program Evaluation: Assessing and Improving Planning, Implementation and Effectiveness*. Sage. approach to program evaluation seeks to provide practical and specific advice for systematically identifying stakeholders' needs, selecting evaluation options and putting a selected approach into practice.

according to Noblit and Hare (1988, p. 17), because in their experience "comparison became aggregating; holism became analysis; etic (the imposition of an outside frame of reference) became preferred over emic; and history became confounding." The background of having conducted the empirical research and with personal connections to each study enables me to understand the contexts to help avoid aggregation. The same depth of knowledge is not available when picking studies off-the-shelf without additional context research. The seven phases of meta-ethnography are also transparent and replicable. All phases and decisions around synthesis can be written-up and interrogated by another researcher working from the same studies.

Meta-ethnography is also well suited to synthesising knowledge from my studies on evaluation because its approach is consistent with resolving social problems. A core element of meta-ethnography and many evaluation approaches is unpacking, theorising and supporting the resolution of social problems. Meta-ethnography studies have almost exclusively focused on social problems in health, education and social care policy and programmes (Britten et al., 2002; Campbell et al., 2011; France et al., 2014; Pilkington, 2018). Similarly, Shadish *et al* (1991), in their classic review of seven evaluation theorists, identify that a core issue for evaluation theory are contributions to resolving social problems. More recently, authors in Hopson and Cram's (2018) edited volume describe evaluation approaches that seek to tackle social and environmental problems in complex ecologies and provide a strategic approach to engage accountability mechanisms.

4.3 Selection Criteria for the Meta-Ethnography

The following section describes the selection and coherence of five of my articles and chapters chosen for the meta-ethnography from a total portfolio of nine. The selection of articles and chapters was based upon six criteria that were informed by the objective of the study and the requirements of a meta-ethnography approach. These are outlined below. In applying these criteria for selection, the nine studies were re-read twice in the selection process, given prior familiarity.

Table 2: Selection Criteria for Studies to be Included in Meta-Ethnography

No.	Criteria
1	The study provides insight on the interaction between evaluation and accountability mechanisms
2	The publication needs to represent research that includes field work in which theoretical terms are anchored in scientifically measurable or observable events (VandenBos, 2013).
3	The study expands on rich theoretical and practical concepts that "explain social and cultural events" (Noblit and Hare, 1988, p. 7). In the selection of articles and chapters for this study two types of conceptual content were sought: • application of a theory to help explain a phenomenon. • adaptation of an evaluation practice or system in a particular social or cultural setting.

4	The study describes the context. Interpretation in a meta-ethnography requires reasoning to encompass "different social and historical contexts, and differing values, norms, and/or social relations" (Noblit and Hare, 1988, p. 29). The texts need to provide descriptions of the context in which studies took place.
5	The articles and studies represent an amalgam of perspectives of professionals who were engaged in the development of an evaluation or evaluation system, providing an emic perspective on debates.
6	The study was published in books and journals and have gone through a substantive peer review process to assure their quality.

The five selected articles and chapters (Porter, 2011, 2013, 2017; Porter and de Wet, 2009; S. Porter and I. Goldman, 2013), met all of the criteria described above and are discussed in the next section.

One article (Porter and Hawkins, 2019) is referenced and forms part of the portfolio informing this thesis, but is not included in the meta-ethnography. The article did not conduct an empirical field work and is not grounded in a particular context (points 2 and 4 in Table 1). The article can be seen as an earlier attempt at synthesis that brings coherence to developing evaluation systems reflecting on the two authors' experience. The article contains pertinent evidence and reflection which are helpful to forming the arguments in this thesis.[11]

4.4 The Coherence of the Five Selected Case Studies

The selected five articles and chapters (Porter, 2011, 2013, 2017; Porter and de Wet, 2009; S. Porter and I. Goldman, 2013), present arguments based on 14 underlying case studies.[12] Each of the 14 case studies sought to understand how and why different types of evaluation systems function through qualitative discussions.[13] Evaluation systems implement a range of evaluation products and streams of data to provide information on overarching goals for an organisation. The systems discussed in the articles were implemented in non-governmental and governmental organisations and often incorporate monitoring as a form of evaluative enquiry.

The articles though largely authored or co-authored by me, nonetheless, provide a polyvocal explanation of professionals writing, thinking, and support to the development of African

[11] The three articles and chapters that were not included in the meta ethnography and are not widely referenced in this work are: Plaatjies and Porter (2011) and Porter (2013) as neither provide insights on the relationship between accountability and evaluation or are empirical; and Porter (2016), is an earlier version of Porter (2017), with the more recent version being preferred.

[12] Three cases are described completely in single articles in Porter(2011), Porter (2013) and Porter and De Wet (2009), six studies are provided in an edited report Porter et al (2012), which were then summarised in Porter and Goldman (2013). Finally, five case studies Adams et al (2013), Alemu and Latib (2013), Porter and Mulenga 2013, Porter and Gasana (2013) and Kumwenda and Latib (2013), were summarised in Porter (2017)

[13] Note that evaluation system is used here as an overarching term for planning, monitoring and evaluation systems (PME) and monitoring and evaluation system (MandE), this is consistent with Scriven's Scriven, M. (1991). *Evaluation Thesaurus* (4th ed. ed.). Sage. definition of evaluation as "a key analytical process in all disciplined and practical endeavours".

monitoring and evaluation systems; representing an ethnographic emic perspective. Each study explores different conceptual frameworks, highlighting challenges faced, and suggesting solutions. The articles and embedded case studies are not a single research project, with funding or a research agenda, rather they provide thinking and ongoing reflection over a ten-year period. The thinking represented has a collective base. The underlying 14 cases are based upon the work of multiple authors (Adams et al., 2013; Alemu and Latib, 2013; Kumwenda and Latib, 2013; Porter, 2011, 2013; Porter and de Wet, 2009; Porter et al., 2012; Porter and Gasana, 2013; Porter and Mulenga, 2013). My role in each case study was varied, sometimes I was co-author, in others I was the overall study manager. The five articles represent summaries of the case studies that were developed through written and inter-personal deliberations with international development practitioners in consultancy, government, non-governmental organisations (NGO) and academia.

The five studies selected are insightful for synthesis because they apply different conceptual frames aligned in their intent to support the aims of international development policy, with an ethic that prioritises development as the expansion of freedom (Sen, 1999). The meta-ethnography provides an overview of how these conceptual frameworks are mutually reinforcing in conducting evaluation that affects accountability mechanisms. Each study reflects on progressively larger political environments in grappling with how evaluation systems can be developed. In the articles the accountability mechanisms and evaluation processes are enabled or constrained by the political systems they work within. There is a chronology to the levels and types of systems grappled with in the article that reflects the development of my career. The below sub-sections describes the studies, the type of evaluation the paper is about, the type of accountability mechanism it reflects on drawing on Table 1 presented earlier and Bovens (2007), the main actors and how they encounter the interaction between accountability and evaluation.

The selected studies consistently reflect upon how evaluation can affect accountability mechanisms. A useful attribute of the studies is that they have been undertaken across different types of evaluation systems in non-governmental organisations and governments. The cases are presented chronologically, representing how my own career moved from community development to working with donors to develop thinking on supporting national systems. By interpreting and synthesising the five studies identified here, we can test the extent to which a more encompassing interpretation of how evaluation affects accountability mechanisms can be developed.

Porter and de Wet (2009). Who will guard the guardians? Amartya Sen's contribution to development evaluation. Development in Practice, 19(3), 288-299.

The article Guarding the Guardians was completed drawing on research undertaken in Zambia and South Africa for an HIV and AIDS programme funded and supported by an NGO in Southern Africa within a community-based organizational (CBO) setting. The article explores the application of Sen's Capability Approach (1999; 2005) in the development of a monitoring and evaluation system. The monitoring and evaluation system, which I implemented, integrated the emphasis on deliberation, public reasoning and participation in the assessment of change from

Sen's writings. The implementation of the evaluation occurred with care-workers who reflected on their work and the successes and challenges in their work within a multidimensional framework (Alkire, 2005). The formal accountability mechanism connected to the evaluation provided information within the NGO and to donors in an administrative manner focused on project performance with reporting conducted hierarchically along a vertical chain from CBO to NGO to donor.

The main finding of the case is that participants valued the process of evaluation because it opened a space for them to influence the work of the implementing NGO, in a way holding them to account. Porter and de Wet (2009) describe how the adapted evaluation processes incorporate accountability mechanisms based on an ethic focused on expanding freedoms through deliberations with community organisations. As a result, the case claims that accountability was also expressed in a professional horizontal collective manner in additional to main vertical streams. The case also discusses how different donors pressurise community organisations through accountability requirements driven by their politics.

Porter, S. (2011). ''Helping' as an Evaluation Capacity Development Strategy in South Africa'. In R. C. Rist, M-H. Boily, and F. Martin (Eds.), *Influencing change: Building Evaluation Capacity to Strengthen Governance.* Washington, DC: World Bank.

The publication 'Recognising Helping' originates from the development of a monitoring system for a President's Emergency Plan for AIDS Relief (PEPFAR) funded project that supported orphans and other vulnerable children in South Africa. On this project I worked as the monitoring and evaluation advisor. The case provides a retrospective study of the process of establishing a monitoring system to meet USAID requirements for valid, reliable, precise and timely information. The accountability mechanism was political and hierarchical, where upon the US congress set targets for PEPFAR and individual NGOs and CBOs had to feed in performance information that were aggregated along a vertical the chain of reporting. The main actors in the case are the staff of the NGO and CBO partners involved in implementing the system.

The study applied a middle-range theory for supporting organisational development (Schein, 2009). The main finding of this study was that the theory of Helping is a useful frame to develop monitoring systems by mediating and bridging between different demands for accountability. Porter (2011) discusses how processes were developed so that donor requirements were helpful also to CBO accountability mechanisms in developing ongoing management discussions and in supporting the delivery of services by community organisations. The ethic of the system focused on improving the rights and capability of children affected and infected with HIV and AIDS.

Porter, S., and Goldman, I. (2013). 'A Growing Demand for Monitoring and Evaluation in Africa'. *African Evaluation Journal, 1*(1), 9 pages. Available from: https://doi.org/10.4102/aej.v1i1.25

The article provides a synthesis of six underlying case studies of governmental Monitoring and Evaluation systems in Africa (Porter et al., 2012). I co-commissioned these studies with the co-

author and managed their implementation. The underlying case studies were written in order to feed into a reflective practitioner workshop between government-led monitoring and evaluation systems from across Africa. The article develops the concept of demand for evaluation.

The article highlights how demand for evidence is connected to an increasingly active citizenry seeking to expand their opportunities. The government monitoring and evaluation systems work within a political hierarchical chain of accountability mechanisms reviewing performance against given benchmarks often within a national development plan. In the article an active citizenry is proposed as a trigger for policy makers requesting evidence, especially in the executive branch of government. And so, responsiveness to citizens can be enhanced through these systems. The main finding of this synthesis is that monitoring overshadows evaluation in governmental systems. In tandem, in some countries there were emerging evaluation systems in place to generate information independently of donors, which feed into national accountability mechanisms.

Porter, S. (2013). 'A Change of road for a rights-based approach? A case study of mobile technology in KwaZulu Natal, South Africa'. In P. Gready and W. Vandenhole. (Eds.), *Human rights and development in the new Millennium: Towards a theory of change*. London: Routledge.
The chapter 'A Change of Road for the Rights-based Approach?' describes an effort to coordinate the work of community health workers with the government health system using a mobile technology system to share information, aiming to address poverty guided by a rights-based approach. I helped to conceptualise and develop this project based on my role as a monitoring and evaluation advisor in the above-mentioned PEPFAR funded project.

In this case, the governmental bureaucratic accountability mechanisms operate in a strict political hierarchy, removing discretion from actors lower in the system and reinforcing information disconnects. The accountability focuses upwards and inwards, rather than to people. The rights-based approach, meanwhile, has little relevance to the day-to-day delivery of services. The chapter suggests developing the concept of joint-production (Clark et al., 2006; Packard Foundation, 2010) to adapt the rights-based approach, enhance the accessibility of the rights-based approach, and also provide a way of working with government accountability mechanisms to strengthen their link to citizens on shared information. The main actors in the paper are staff of NGOs, CBOs and government health service providers.

Porter, S. (2017). 'A Framework for Identifying Entry Points for International Development Evaluation to Enable Responsive Government Policy'. In D. Podems (Ed.), *Democratic Evaluation and Democracy: Exploring the Reality*. Charlotte, NC: Information Age Publishing.
The chapter focuses on how evaluation can enable responsive government policy through a synthesis of five case studies of country monitoring and evaluation systems using a principal-agent perspective (Lindberg, 2013; Stiglitz, 2002). The studies and synthesis were funded by DFID to inform their work on developing country monitoring and evaluation systems. I managed

the implementation of these studies and wrote two of the studies (Porter and Gasana, 2013; Porter and Mulenga, 2013).

The chapter highlights how the political economy of monitoring and evaluation systems within a country enables and constrains the use of evidence. The chapter argues that the political economy defines the spaces in which evidence is likely to be treated as useful at all levels of the system, from the community organisations to cabinet meetings. The chapter identifies that for the evaluation system to enable responsiveness to poor people, you need to work with an understanding of the politics to define likely entry-points to assist changes in practice. The chapter demonstrates that there are a range of types of accountability mechanisms in play in national evaluation systems, although the ostensible logic might be vertical, there are spaces and opportunities to frame how accountability mechanisms function perhaps in horizontal, diagonal and indeed citizen facing manner. The main actors considered in this article includes government, evaluation professionals, academics and think tanks, NGOs and politicians.

5. Undertaking the Meta-ethnography the Method for Reading, Relating and Translating the Studies - Phase 3-6

The following section provides the methodology and illustrations of the process for phases 3-6 of the meta-ethnography. Each of the phases in this section builds upon the literature review - helping to get started (phase 1) -, and the objectives, questions and study selection that defined what is relevant (phase 2) for this study. The following section describes each phase of the meta-ethnography that involved the process of reading, juxtaposition, translation, interpretation and synthesis – phases 3-6 of the meta-ethnography. These phases take place simultaneously.

In reviewing the studies, two approaches have been undertaken: Holistic reading and dynamic reading. Holistic reading reviews the whole texts for the overall arguments and ideas developed. Dynamic reading, meanwhile, reviews texts for topics, and seeking to represent their different meanings. Throughout the process of synthesis (phases 3 to 6) dynamic and holistic reading of the selected texts took place.

5.1 Reading and Rereading the Studies - Phase 3

The in depth reading and re-reading of studies was initially undertaken chronologically starting with the earliest article, in alignment with the advice of Atkins *et al* (2008). Chronological reading allows the identification of how topics have emerged over time. This is important as the studies show how my thinking on the theory and practice of evaluation evolved as my career developed. Initially, holistic reading was the dominant way studies were interacted with. As the research advanced, holistic reading diminished and dynamic reading increased.

<u>Holistic Reading</u>
Two rounds of reading the complete studies was undertaken. The context of the study and some of the main issues arising were noted.

<u>Dynamic Reading</u>
Initially, extracts of interest were identified, but there was no reading for individual topics.

5.2 Determining how the Studies are Related - Phase 4

Determining relationships between the studies involves interpretation and juxtaposition. This is the first phase in which dynamic reading occurred. The chronological reading of texts was maintained. Following the advice of France *et al* (2019), specific aspects of the study were targeted in determining how studies are related, as described below.

<u>Holistic Reading</u>
Three rounds of systematic reading of each of the five texts was completed, focused on how accountability and evaluation connect. The first read identified the theory and linked elements that were applied in each study. The second read identified topics related to evaluation practice

and how they interact with accountability mechanisms. The third read drew out further contextual issues defining how accountability mechanisms functioned. Through this process, similar topics within articles were identified and extracted into a separate new analysis for each study. A single topic could have multiple related extracts from different areas of the same study.

Dynamic Reading

After each round of reading individual studies, topics were noted, and the related text extracted. Each cluster of topics was then interpreted and a summary interpretation produced describing the theme arising from the cluster of topics. The process was grounded, but also informed by the issues that emerged during the literature review. The themes that emerged from the topics and their interpretations across all texts were then transcribed into an overall table and juxtaposed. Table 3 below provides an example of the interpretation that generated a theme during this phase.

Table 3: Example of Theme Generation

Theme	**Topics**	**Interpretation and related themes**
Evaluation challenges those who hold power	"We argue that by giving a central position to public discussion, social agitation, and open debate, and understanding development as freedom, a wider array of stakeholders can exercise power in a more open and explicit manner." (Porter and de Wet, 2009, p. 292) "The maxim '*Quis custodiet ipsos custodes*?' (Who will guard the guardians?) (we discuss the identity of the guardians below) is a matter of central importance in political science, and an appropriate principle to guide us in our critical reflections on development practice." (Porter and de Wet, 2009, pp. 297-298) "These questions, regarding the distribution of power in development, if they are addressed by reasoned debate and if they incorporate input from the disadvantaged, can challenge those in power – the guardians." (Porter and de Wet, 2009, p. 298)	Reasoned debate and participation help to generate tensions and challenge those in power. The concept relates to others that focus on affecting agents with power

The process of dynamic reading also helped to support my own reflexivity. In reflecting upon the article or chapters important elements around my role, the origin, context and connections between the studies became apparent. In phase 4 emerging themes were interpreted for each article and similar interpretations were identified and sign-posted across studies.

5.3 Translating the Studies into one another - Phase 5

Translation of studies into one another takes place by identifying how studies are like each other through the refinement of themes into concepts and the identification of analogies between the studies. The development of the concepts took place in a manner that sought to protect the holism of each study by maintaining the key metaphors that arose in the underlying data.

Holistic Reading

The studies were read holistically at the end of the process of interpretation to check that the concepts were analogous to the issues identified in each study.

Dynamic Reading

The overall study question 'How does evaluation affect accountability mechanisms?' was referenced to identify and guide translation from themes into concepts that explain a specific issue. Through this process the interpretations moved beyond the themes identified in the selected texts. By further juxtaposing themes across articles, and checking back against the overall question, new concepts started to emerge in more detail. These concepts were given working titles. A new table was generated that helped to compare and contrast underlying themes and interpretations. In the process of translation, notes on the context and meaning of each case were cross-referenced by maintaining links to the underlying text.

In Noblit and Hare's (1988) terms, reciprocal synthesis took place where "in iterative fashion each study is translated into the terms (metaphors) of the others and vice versa" (Noblit and Hare, 1988, p. 38). In developing concepts (metaphors) a more complete understanding of all studies was achieved, providing an initial overarching synthesis. Through this process, in alignment with Campbell et al (2011), concepts were not stripped of context and meaning through recursive summarisation.

The example below in Table 4 demonstrates how themes from three different studies were translated to develop an overarching concept. Each study articulated a different aspect of how power is held in relationships when implementing an evaluation. The first example highlights tensions generated with those who hold power through evaluation, the second talks to the importance of managing power relations within evaluation, the third highlights ways to manage relationships in an equitable manner.

From the three different themes a new concept emerges that emphasises that evaluation affects accountability when the process seeks to engage agents with power with an understanding of their perspective. In this way we move from a discussion that the tactics of evaluation need to change to improve the way it's used, an issue covered by previous authors (House, 1980; House and Howe, 1999), to a more strategic and substantive engagement whereby an evaluator seeks to perceive an issue from the perspective of those with power.

Table 4: Example of Translation from theme to concept

Article/ Chapter	Theme in paper	Interpretation	Concept
Guarding the Guardians	Evaluation challenges those who hold power	Reasoned debate and participation help to generate tensions and challenge those in power.	Access agents with power to understand their truths/ perspectives within evaluation processes
Responsive Government Policy	Managing power relations	Understanding power and situating evaluation within the relational domain of the evaluation process. This is not only about recognising power but having a systems perspective on how to develop power relations.	
Helping	Relationships between different actors are based on fair and equitable practice for those involved	Relationships integrate the technical and the social, they enable progress to be made on difficult issues and moving at different speeds with the audiences and economics.	

The concepts and their corresponding interpretations were mapped and remapped against the articles, until eventually all concepts could be related to a cluster. Through this process the clusters of concepts were gradually collapsed, generating four overarching concepts. The titles of these concepts were then refined to represent overall translations of the underlying issues and interpretations. In the process the different theoretical starting points (e.g. capability approach, rights-based approach) and circumstantial nuances were maintained in the underlying descriptions.

Through the process concepts emerged at two different levels of interaction and relationship building, as summarised below.

1. In the authorising environment - Navigate the authorising environment based on a goal
2. In the evaluation process - Expand the focus of demand
3. In the evaluation process - Access and respect the agenda of agents with power
4. In the evaluation process - Undertake a shared journey on evaluation quality

The concept of navigating decision-makers based on a goal emerged as predominantly operating in the authorising environment. Accountability mechanisms and evaluation processes operate within overall authorising environments. The authorising environment convenes agents in formal spaces and represents a strategic space of decision-making in Fox's (2015) perspective. An authorising environment links the formal chain of agents that hold accountability for making decisions; without their authorisation resources will not be formally allocated to an intervention (Benington and Moore, 2011). In each of the studies interpreted,

authorising environments are defined; each being context specific, distinct in both their tangible (location and people) and intangible (culture, values and political economy) aspects. In the authorising environment, evaluators need to operate and form relationships in the decision-making system where agents with power meet and set agendas.

Meanwhile three concepts emerged predominantly at an analytical level related to the evaluation process. While many of the technical parameters of the evaluation process have been well defined, the specific guidance around relational processes still forms a more limited part of the literature, only recently being more fully detailed (Hopson and Cram, 2018; Marra, 2018). An evaluation process typically involves the stages of planning, commissioning, undertaking, reporting and follow-up (Hawkins, 2010; Patton, 2008). The evaluation process can be conducted before, during or after implementation – whether for a process, product, program, product or policy (Chen, 2005; Owen, 2007). An evaluation process represents a tactical space where only some degree of change in accountability mechanisms can occur. The process of translation suggested that expanding the focus of demand, accessing and respecting the agenda of agents with power and undertaking a shared journey on evaluation quality all contributed to affecting accountability mechanisms through the instrumental use of evaluation. This goes further than the literature review, which only partially represented how relational connections to accountability mechanisms could affect the use of evaluation.

The four concepts and contexts were then reviewed to understand whether these translations could be elaborated into a line-of-argument synthesis to connect these concepts together to form a middle-range theory (phase 6 of the meta-ethnography).

5.4 Synthesising the Translation - Phase 6

Until now the translations have been developed within channels generating separate clusters of concepts developed. A line-of-argument synthesis goes further by identifying how individual concepts connect, to anticipate how evaluation can affect accountability in other situations. Lines-of-argument synthesis are new interpretations that help to bring together the similarities and contrasts of cases into an overall argument (France et al., 2019). To undertake a line-of-argument synthesis we start to examine the connections between the translations and understand how context affects each. This means analysing each translation seeks to generate interconnections and understand differentiating contextual conditions.

Holistic Reading

The original articles were read again holistically upon summarising the clusters, to check the translations and then overall synthesis. None of the translations was found to be out of alignment or jarring against the original text.

Dynamic Reading

In developing and testing a lines-of-argument synthesis the translations were reviewed against the underlying extracts and narratives. In the review, similarities and differences in emphasis were identified to further nuance understanding of the overall translation. The successive

comparison of similarities and differences in emphasis helped to refine the interpretations and translations. Noting the similarities helped to focus the description to ensure the capture of the main points for each concept. Understanding the different areas of emphasis brought out nuances in how the interpretations worked together. Through this process the development of headline connections between cluster concepts took place.

Similarities and differences were compared in light of the contextual conditions in each study. Maintaining a link from translations to the context in which they occurred distinguishes meta-ethnography from other more aggregative forms of qualitative synthesis (Noblit and Hare, 1988). In contrasting similarities and differences their salience was crossed-referenced against the type of evaluation process, the organizational conditions in which they arose, and the kind of agents involved and their relative power.

In working through each translation in this way, the connections could be understood: how each "fit" and "work" together (Noblit and Hare, 1988, p. 63) . At the broadest level each cluster is about the astute management of relationships as the main pathway to affect agents who exercise power in accountability mechanisms. Magnifying the details of each concept emphasises differentiated aspects of relationships. Building on the examples presented earlier, Table 6 shows how nuance of the overall translations develops through examining similarities and differences.

Table 5: Similarities and Contrasts

Article/ Chapter	**Themes and Topics in paper**	**Similarities**	**Contextual Conditions**	**Contrast with other concepts**
Guarding the Guardians	Evaluation challenges those who hold power – "We argue that by giving a central position to public discussion, social agitation, and open debate…a wider array of stakeholders can exercise power in a more open and explicit manner." (Porter and de Wet, 2009, p. 292)	Emphasis on dialogue and debate in the relational nature of evaluation	Participatory development orientation in organization Conducted at grassroots level physically and psychological distant from authorising environment	Though public debate is highlighted, the concept does not engage with the broader power structures in society.

Responsive Government Policy	Managing power relations – "While principals transfer power to an agent, there are issues with who monitors the transfer of power. In countries where development aid is delivered, power is often centralised in a strong executive or elite group. Knowing how the elite group operates helps to identify where accountability mechanisms work for private, rather than public interests." (Porter, 2017, p. 25)	Emphasis on the different relationships needed.	Case studies took place in context where elite interests and different forms of rent seeking dominated the logic of development.	There is a wider system of power and relationships beyond an evaluation study. Different agents need to be heard differently
Helping	Relationships between different actors are based on fair and equitable practice for those involved "Helping encompasses developing horizontal relationship between the helper and the helped rather than vertical teacher and student relations." (Porter, 2011, p. 157)	Relational and 'horizontal' relationships - implied is the importance of dialogue and debate.	Donor determined development provided pathways for NGO staff to develop career and CBOs to garner resources for local elite and local needs.	The approach of Helping enables the evaluator to connect to different perspectives.

Having developed and cross-referenced all the similarities, contrasts and then elaborated the overall description of each, I undertook a cross-check of my own articles and chapters to understand whether there were further insights that could be related to the overarching translations. For example, issues around capacity were not initially picked-up across all readings; through further iteration and the use of word search, new extracts around capacity development were identified that reinforced the overall synthesis.

By developing the similarities and contrasts we move beyond describing the translation to affirming new concepts and identifying the core elements of a middle-range theory that includes:

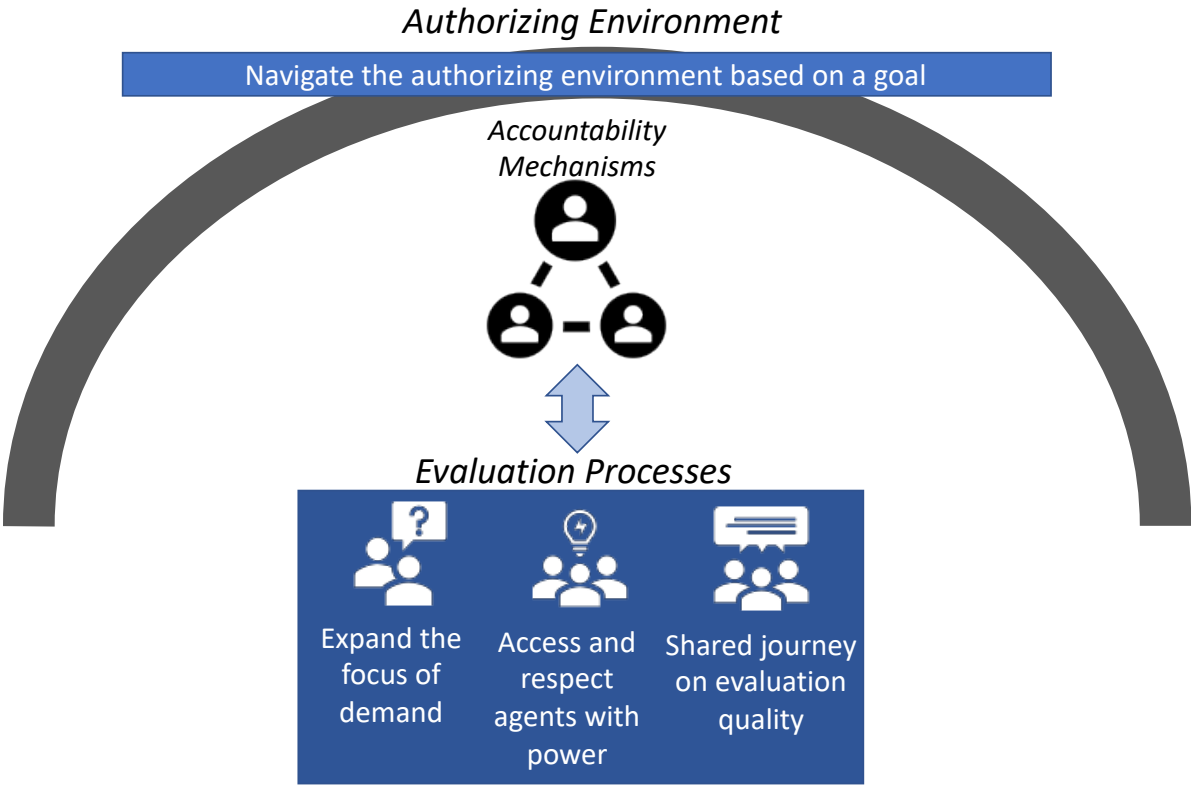
- *Where* – The most salient places where each concept affects accountability mechanisms
- *Why* – The importance for understanding how evaluation affects accountability
- *What* – The targeted areas of practice of change through this concept
- *How* – The practical approaches through which the concept comes into operation

The synthesis generated does not specify the agents targeted to form relationships (who). At a broad level, agents with power should be engaged with, but these agents vary depending on the political economy of the context. For example, in a non-governmental organisation, agents were engaged in the donor community and in the local geographic communities. Whereas when operating at the level of the government system, agents potentially engaged include parliamentarians and managers in the public service. The core elements of a middle range theory are elaborated upon in the next section that represents the adopted line-of-argument synthesis.

6. Expressing the Synthesis - Phase 7:

The main message of the synthesis is that accountability mechanisms are affected by evaluation through the combined interaction with strategic and tactical processes that relate to agents in accountability. We can anticipate that accountability mechanisms are affected when evaluators work through four overarching relational concepts. Relational practice pays attention to how the ability to work and communicate with others builds or crumbles through speech, acts, behaviours, interactions, affected by contexts and significant power differences that can dominate others (Hopson and Cram, 2018). Strategically the middle range theory prioritizes navigating the authorising environment in order to develop partnerships that provide leadership and exercise power towards a common goal. Three tactics implemented in the process of evaluation support strategic change in the authorising environment. When evaluators work towards a common goal in an accountability mechanism's authorising environment and activate the three tactics in the process of evaluation, the logic of decision-making adapts. Figure 5 illustrates and summarises the concepts.

Figure 5: Four Concepts through which Evaluation affects Accountability Mechanisms



In each of the studies within the meta-ethnography, elements of all of these concepts are activated or identified as important. The following section dives into the detail of the concepts starting with the authorising environment, the strategic domain for affecting accountability mechanisms. Discussion of the location of concepts, why they are important, and what

processes change as a result of their implementation are presented for each concept. How to activate concepts cuts across the studies and discussed in regard to their different approaches.

6.1 The Authorising Environment - Evaluators need to operate and form relationships in the decision-making system where agents with power meet and set agendas.

Evaluation theory and evidence-based practice often focuses on delivering an evaluation or making tactical changes to processes of interaction as demonstrated in the literature review. Many evaluators, however, spend time placing themselves, working in and relating to authorising environments, where the capability for power to be exercised in a manner that dominates other decision-making structure exists. These examples show that authorising environments provides a strategic location to affect accountability mechanisms. In working in the authorising environment evaluators come into closer contact with political decision-makers and can influence the criteria applied in accountability mechanisms, thereby providing a strategic point for influencing change (Fox, 2015). Similar arguments about the importance of the authorising environment to evaluation are also made in the context of governments by Lopez-Acevedo et al, (2012) and Picciotto and Wiesner (1998); and by Holvoet and Dewachter (2013) in relation to evaluation associations.

6.2 Navigating the Authorising Environment based on a Goal

To affect accountability mechanisms evaluators need to develop relationships across an authorising environment with agents who share common goals that may include social betterment (Mark et al., 1999), oversight (Chelimsky, 2006), transformation (Mertens, 2009) or performance (Rossi et al., 2018). Evaluators can be assisted in understanding how power operates in visible, hidden and invisible forms to enable or constrain opportunities. Across all the studies the ethical orientation in terms of Sen's (1999) conception of expanding freedom remains the consistent frame to navigate in the authorising environment.

Why is navigating the authorising environment important?

When evaluators develop a vision of change with others in the authorising environment then they are able to affect how criteria are formed and executed across a range of accountability mechanisms. In each of the studies interpreted, working across sections of the authorising environment was important to refine agreement on the criteria and standards deemed to be relevant, the first two steps in the logic of evaluation. This occurred whether in the implementation of a monitoring system or in identifying specific entry-points for improved evidence-based practice.

Authorising environments work in a network fashion (Sørensen and Torfing, 2016). Therefore, the development of a shared vision in an authorising environment requires work across networks of agents. The following quote from Porter (2013, p. 237) reinforces that there are real world consequences of not consolidating criteria of success across a network:

> Over a period of two years, some hospital equipment awaiting collection and transport to the local hospital was left outside. It was once reasonably usable, but slowly

> disintegrated over time. Where beds once were, now rusty health hazards stood in a clinic's yard. The community organisation that witnessed this would not challenge the hospital for fear of repercussions. The local clinic attempted to intervene but the hospital, to which the clinic reported, did not respond. (Porter, 2013, p. 237)

Navigating decision-makers is insufficient in itself . To affect accountability mechanisms across authorising environments, evaluators need to navigate decision-making based on an ethically informed goal, in order to work in a professionally proper manner and connect appropriately those with the capability to exercise power.[14] For evaluation to affect accountability there needs to be a vision, a set of values and a philosophy that guides implementation and the development of evaluation systems that are attuned to the goals of agents in the authorising. Without a goal that guides the development of relationships, the change that evaluators are seeking to effect will remain unclear and they will not connect with the visible and hidden agendas of agents or navigate invisible systems of power.

Across the studies the expansion of freedom based upon Sen's capability approach was an explicit or underlying goal that informed interactions with the authorising environment. Porter and de Wet (2009) provides the most explicit connection arguing that taking a view of development as freedom enabled care workers to hold the implementing agency to account, in a manner that resonated with donors. Porter and Goldman (2013) and Porter (2017) argue that the goal for monitoring and evaluation systems should be enhanced responsiveness to citizens and in this they are informed by a perspective of Sen. The rights-based approach of Porter (2013) closely links to the arguments of Sen in terms of public reasoning and assessing the capability of individuals. Porter (2011) in describing Helping outlines an approach to develop and undertake reasoned debate, in a similar vein to Sen.

What changes in evaluation through navigating the authorising environment?

An evaluation process by itself is unlikely to shift an accountability mechanism. By navigating the authorising environment evaluators can help shape the way accountability mechanisms interact. In the studies that fed into the meta-ethnography, descriptions of the authorising environment emphasise the importance of identifying power dynamics in visible, hidden or invisible forms.

To navigate the authorising environment, evaluators and those managing evaluation systems maintain a focus on working in the broader politics of their environment, sometimes called the political economy. To navigate the politics, evaluators need to implement tools to analyse the roles of agents in demanding evidence and developing relationships with those who have power to set the policy agenda and see the importance of evidence in improving practice.

[14] In the evaluation standards developed by Yarborough *et al* (2011) working in a proper manner is termed proprietary and entails amongst other standards a responsive and inclusive orientation and human rights and respect.

Porter (2017, p. 21) develops the concept of the authorising environment within an evaluation setting by working through principal and agent linkages and . Porter (2017) highlights that agents become principals when power is assigned to a specified set of agents, for example, when a citizen transfers power to an agent (politicians) to act on his/her behalf through elections, or through guaranteeing security where there is a history of violence. Principals in the state apparatus broadly define the prevalent configuration of the authorising environment and are the demand side of evaluation. In this role they have the capability to exercise power through the state apparatus, stop issues from emerging and unconsciously maintain cultures of domination. Porter (2017) further outlines a range of tools to reflect on the different roles and motivations played by principals in an evaluation system. These are discussed further in section 6.7.

To form relationships evaluators, need also to change how they work to improve information flows between stakeholders in the authorising environment. There are asymmetries of information between those governing and those governed, and just as markets strive to overcome asymmetries of information, evaluators need to look for ways by which the scope for asymmetries of information between political agents can be limited and their consequences mitigated (Porter, 2013) in order to generate spaces for the contestation of power or surface invisible unconscious domination. For example, although donor demand for evaluation can usurp country evidence requirements, the studies show that it is too simplistic to consider donor demand for evaluation as narrowly focused upon their own interests.

As Porter (2017) highlights, on occasions where government takes an interest, donors' demand for evidence can lead to shifts in implementation. For example, in Malawi the government used the recommendations from an evaluation commissioned by donors of a cash transfer program to expand the program. Evaluators and managers of evaluation systems need to ensure their practice helps different agents to see complementary roles based on their respective capabilities to exercise power.

Porter and Goldman (2013, p. 2), meanwhile provide a dynamic view of how an authorising environment functions between different interests related to a results-based orientation to government. They argue that demand as an active expression of power is shaped by both rent-seeking and performance-focused groups. Yet, this expression of power is formed within barely documented processes within countries where interests are negotiated. Consequently, in navigating the authoring environment evaluators can identify strategic opportunities for taking forward a results-orientated reform agenda, using evidence to support improvements in delivery.

Porter (2009; 2011) describe how NGOs work in a multi-faceted authorising environment where organisational management, community partners and donors operate together. In NGOs donor demands in the authorising environment can have both positive and negative effects. Porter and de Wet (2009, p. 296) highlight the effects of the broader international development authorising environment as well as the organisational management processes in dominating how the reporting of development occurs. The authors argue that the overarching environment

restricts what can be imagined, creating an invisible force constraining action. Porter and de Wet (2009) illustrate how donor demands hamper the assessment of change related to HIV and AIDS responses. This leads to an assessment where evaluators seek alternative types of accountability relationship outside of the hierarchy of reporting and in Porter and de Wet (2009) a more horizontal evaluation practice.

In contrast, in developing a project monitoring and evaluation system, Porter (2011, p. 165) highlights that in responding to the exercise of power through donor demands enabled more positive discussions of performance in a variety of authorising spaces:

> "First, the [donor] data quality audit revealed an immediate improvement in the data...these improvements satisfied the demand for evaluation at the AMREF level. The evaluation team was able to enter into detailed discussions with partners about the way the project operates, how their staff work, and how the management support decisions—all based on up-to-date, reliable evidence."

Porter (2011) illustrates that by considering the donor driven authorising environment evaluators can identify areas to expand evaluation practice within their organisation.

6.3 The Evaluation Process – Focus on Interrelationships that Substantively Affect Accountability Mechanisms

The synthesis shows that evaluators form relationships and affect accountability mechanisms in a tactical manner, namely: expand the focus of demand; access the agenda of agents with power; and undertake a shared journey on evaluation quality. All three of these concepts interact with power tactically, in that they are bounded within an evaluation process and work mainly on changing the terms of information provision (Fox, 2015). Implementation of these concepts supports evaluators working in the authorising environment in that work towards goals is grounded in actual processes of evaluation. This does not mean, however, that these concepts can be directly applied in the evaluation processes. The empirical evidence in the meta-ethnography only highlighted their application in evaluation processes, rather than the authorising environment.

6.4 Expanding the Focus of Demand

Evaluators often face constraints in the active demand for evidence. To affect accountability mechanisms, evaluators need to work with agents who hold power to dominate others and expand their range of demands for evidence, aligned to the goal they are seeking to achieve.

Why is expanding the focus of demand important?

Even with evaluators effectively navigating the authorising environment it is likely that actual evaluation processes will remain constrained by forces of power beyond the influence of the evaluator. The main constraining factor identified in the articles and chapters was the structure of demand during evaluation processes, which provides the visible structuring of power

relations. A demand for evaluation arises when "decision makers require or are influenced to ask for evidence" (Porter, 2017, p. 21).

Porter (2011, p. 155) notes that:

> "the effectiveness of evaluation is defined by how the demand is structured for improved results and performance. When short-term goals are the driving force, the direction is inappropriate and rewards are perverse; independent, credible, and useful evaluation capacity cannot be developed."

The studies identify that the information base prescribed by accountability mechanisms often narrowly focuses on fulfilling limited informational requirements. In particular, in monitoring systems the results-based notions that were applied, generated in francophone contexts "perverse incentives that reinforce upward compliance and contrôle to the detriment of more developmental uses of monitoring and evaluation evidence." (S. Porter and Goldman, 2013, p. 2). Meanwhile Porter and de Wet (2009) noted that the manner in which classic control-group methodologies had been prescribed by law in the USA, generated a demand that excluded other lines of enquiry or knowledge. Expanding the focus of demand is important because it affects the information deemed to be relevant during an evaluation process.

Evaluators, even in limited ways, should look to find ways to expand the focus of demand, working through constraints in the evaluation process by identifying spaces of contestation or unconscious exercises of power that are divergent from the aims of international development (see 6.7 for discussion of approaches that enable the surfacing contestation). The evaluation process will then have the potential to instil, reinforce and challenge current demands. In an evaluation process, attempts to shift demands can, for example, focus upon the questions being answered. By contrast, in the authorising environment the evaluator seeks to shape the overall system of demand by navigating the interests of different agents rather than answer specific questions.

What changes through expanding the focus of demand?

Expanding the focus of demand means that evaluation processes nurture incipient demand for evidence. Nurturing demand also recognises constraints and instils a sense of realism about the extent of change possible within the accountability mechanism. Evaluations are commissioned by a type of accountability mechanism often working in a vertical, performance orientated manner, whether as a funded independent unit or as consultants. As a result, when conducting an evaluation, certain choices are bounded and targeting the expansion of demand means being thoughtful about where to spend your relational capital.

Porter and Goldman (2013), do not suggest revolutionary action like replacing the system, but rather propose working with agents aligned to shared goals who hold a capability to exercise power through evaluation and have yet to do so. For example, the authors suggest working with legislatures and donors in reporting processes (Porter and Goldman, 2013), which will help to surface potentially hidden contestation of agendas. Porter (2013) argues that to expand

the focus of demand requires greater inclusion and deliberation in deciding what information is relevant in the form of processes of joint-production. The arguments on joint-production resonate with authors in the evaluation literature who emphasise the importance of participation and shared decision making in evaluation processes (Patton, 2010; Ryan et al., 1996; Van Der Meer and Edelenbos, 2006; Zapico-Goñi, 2007). Operating in a constrained manner, focusing on how to include people in different ways in the evaluation process, recognises the limited formal power of the evaluator and the need for allies in expanding the focus of demand.

6.5 Access and Respect the Agenda of Agents with Power

Evaluations often neatly categorise people either as respondents to the evaluation questions or as agents they report to. Accountability mechanisms are in practice not so neatly delineated, with people straddling both groups. To affect accountability mechanisms, evaluators need to understand what interest groups with power want out of a situation by accessing and reporting to and against their perspectives through the evaluation.

Why is accessing and respecting the agenda of agents with power important?

Unless an evaluator can work across and engage in reasoned debate with different agents who hold power, they cannot access the different perspectives that agents hold in accountability mechanisms. Agents with power are not just those with formal authority within an evaluation; they also include people who are being targeted by an intervention and exercise their own agency in an evaluation to push their interests. Engaging with agents is not about ceding an evaluator's independence; but rather about understanding the agents' perspectives in order to speak truth to those with the power to dominate others in a way that can be heard. This concept overlaps with expanding the focus of demand and developing a shared journey on evaluation quality by enabling evaluators to work in a way that better communicates with agents. The concept of accessing and respecting the agenda of agents with power aligns with culturally competent evaluation that recognises that evaluators need to move beyond their own narrow assumptions and work with "diverse sociocultural perspectives and experiences" (Hopson, 2003, p. 1).

Porter and de Wet (2009, p. 289) provide a metaphor that helpfully frames the issue in relation to the exercise of power, the maxim "'*Quis custodiet ipsos custodes*?' (Who will guard the guardians?)" In asking who will guard the guardians we are reflecting that the potential to exercise of power also needs to come with its constraint. Accountability mechanisms exercise and constrain power through mandates that define criteria, agreed standards, measurement and synthesis of the potential deployment of sanctions. An evaluator that affects accountability mechanisms will understand both the exercise of power and its constraints in undertaking evaluations.

What changes through accessing the agenda of agents?

Accessing different agents to understand their agendas changes the conduct of evaluation by connecting to, representing and showing respect for differing perspectives. This approach to

accessing the agendas of agents is aligned in the chapters and articles interpreted to an ethic that focuses attention on expanding freedom is developed. Within accessing the agenda of agents, particular attention should be paid to hidden and invisible power as well as more formalised roles that constrain the realisation of freedom.

An important area where the evaluation process changes in accessing the agenda of agents is when evaluators seek to undertake inclusive discussion during an evaluation. For example, through the application of the Action Learning process, Porter and de Wet (2009, p. 295) sought to access alternative definitions of success and support mutual accountability. Through this process the care givers in the evaluation were able to articulate their own view of success:

> "A statement from the Zambian participants movingly summarises the majority of recommendations of both groups of care givers during the focus groups: 'First you take care of the care giver, then you take care of the client'."

The point highlights that the care givers themselves are active agents using the evaluation process to further their own agenda and contest the power of others. By understanding this perspective, the evaluation process in reporting and feeding back was better able to acknowledge their perspective and prioritise it in regard to other agendas. In this case, the feedback from the care workers reinforced an existing orientation of the programme to provide further support and to develop the capacities of community care givers, as a niche area, given that many other organisations were focusing on the quality of care.[15]

Porter (2017), demonstrates the importance of accessing the agendas of agents in a systemic manner by applying a principal-agent framework (Lindberg, 2013; Stiglitz, 2002). The evaluation changes as the evaluators seeks to connect to the roles of principals, intermediaries and agents. For example, the chapter further develops the segmentation of agents with the idea that there are intermediaries who work in-between principals, who demand evaluation, and agents that undertake evaluation processes and activities.

Porter (2011) provides an example of where insufficient examination of different interests led to gaps in the influence of an evaluation. In analysing the process of developing a monitoring system for community care organisations the theory of Helping developed by Schein (2009) was applied. Porter (2011, p. 165) noted that in focusing on the care workers in developing the monitoring system, other groups such as managers of community organisations became less involved. Had managers been more fully involved, it is possible that they could have better accessed organisational power structures and better supported information use.

Recognising that different agents are constraining as well as exercising power (guarding guardians) requires that the evaluator, in accordance with the advice of Sen (Gasper, 2002), respects that agents are authors of their own destiny in exercising their judgement. The evaluator will seek to be influential through their findings, by expanding the criteria and by

[15] The orientation to care workers is demonstrated through VSO-RAISA's (2007) Conference report

developing evaluative thinking (e.g. surfacing contestation, providing counter-factual scenarios and identifying unconscious reasoning). But they accept, humbly, that agents are exercising power and, in the end, own the decisions they make.

As the evaluation process changes to pay attention to the agenda of agents, so improved connections to accountability mechanisms are developed, which allows for the prioritisation of findings aligned to common goals. Respecting that people will use evidence to be authors of their own density is an important element of the change, a point that resonates with practices of democratic evaluation (Kushner, 2002). By understanding different agendas, whether those with more or less influence, an accountability mechanism can be affected.

6.6 Undertake a Shared Journey on Evaluation Quality - Evaluation Capacity is Developed with Interests through the Process of Evaluation

Agents in accountability mechanisms need to be assured of the quality of evaluations if they are to use them (Lopez-Acevedo et al, 2012). The development of capacities through evaluation processes required the definition of appropriate evaluation quality for the context (King, 2007; Preskill and Boyle, 2008; Taylor-Powell and Boyd, 2008). The definition of evaluation quality in a specific context is a shared endeavour requiring the participation of a range of stakeholders (Hawkins, 2010).

Why is undertaking a shared journey of evaluation quality important?

Evaluation capacity enables agents in accountability mechanisms to debate and request appropriate forms and monitor the quality of evaluation. When agents can identify the values that they want to assess, the issues they would like to question and the range of methods that are used, they will be able to confidently enter into conversations with evaluators. Therefore, they will be able to better manage the quality of evaluations focusing on an agreed goal. Practical capacity will only be really accrued during an evaluation process, though conceptual knowledge can be developed in training courses. Evaluators have a role to play in developing capacity for evaluation during the process of the evaluation.

A wide variety of texts have made arguments about what constitutes appropriate methodological and process quality in evaluation (Davidson, 2005; Duflo and Banerjee, 2017; Patton, 2008; Shadish et al., 1991). These methods texts do not often bring out that quality needs to be developed in accordance with local demands for evidence in order for the evaluation to be influential; this issue is more deeply covered in the evaluation capacity development literature (Boyle et al., 1999; Lopez-Acevedo et al., 2012; Mackay, 2007). Porter and Hawkins (2019, p. 94), for example, argue that:

> "standards of appropriate methodological and process rigor are… put into operation for specific organisational environments in developing an evaluation system. Different organisational evaluation systems often define their own typologies of evaluation, interpretations of quality standards, methods, and quality review mechanisms depending on their own institutional demand for evaluation."

In practice there is information asymmetry between evaluators and accountability forums; the expertise of evaluators can be a barrier to developing quality. By developing capacity, evaluation is not seen as the preserve of a limited number of experts, everybody can situate and see themselves as an evaluator. As numerous agents situate themselves in an evaluation process, so the journey on evaluation quality becomes one shared with the evaluator, forming a common language. This also helps to reduce risk in the process of evaluation, since commissioners and evaluators will be able to jointly clarify expectations of success and identify how the process interacts with their own positions of power (Porter, 2013).

What changes through undertaking a shared journey on evaluation quality?

Undertaking a shared journey on evaluation quality leads to an increased focus on developing capacities of agents who manage and commission evaluations. A challenge in developing capacity is that evaluative thinking that supports reasoned contestation at both the individual and collective levels need to be developed. Buckley et al (2015, p. 376) argue that evaluative thinking at an individual level is:

> "critical thinking applied in the context of evaluation, motivated by an attitude of inquisitiveness and a belief in the value of evidence, that involves identifying assumptions, posing thoughtful questions, pursuing deeper understanding through reflection and perspective taking, and informing decisions in preparation for action."

Schwandt (2018) added to this argument that evaluative thinking works collectively by engaging people in deliberate collaborative processes that help to understand how different interests frame or bound solutions.

Porter (2013) demonstrates that capacity for evaluative thinking needs to be developed throughout organisations in a manner informed by power differentials, and not just in evaluations. For example, support to project management and information processes supported community organisations to engage with developing rights-orientated evaluation capacities that could be applied in shaping decision-making with duty bearers in the state (Porter, 2013). Developing only the technical resources on the supply-side of evaluation will not lead to the improved use of evaluation and underlines the importance of capacity development of agents who demand evaluation.

6.7 How do the Four Concepts come into Operation?

Evaluators need tools and ways of working to affect agents in the authorising environment and the evaluation process, in order to better reach the goal they are working towards. Putting a concept into action requires tools and ways of working that can be applied in different contexts. Without elaboration of *how* concepts operate, they remain elevated and difficult to put into practice as a middle-range theory. This section grounds each concept by showing how they work together to affect accountability mechanisms.

The four concepts come into operation through approaches that provide tools and ways of working transferable between contexts. Each study the synthesis applied an approach: Porter and de Wet (2009) explicitly apply Sen's capability approach (Sen, 1999); Porter and Goldman (2013) analyse the structure of demand; Porter (2017) presents a political economy analysis approach; Porter (2011) adapts a middle-range theory of Helping from management consultancy; and Porter (2013) develops the joint production.

The synthesis anticipates that a blend of approaches will most effectively support the four concepts to affect accountability mechanisms. In presenting how the approaches come into operation, illustrative examples are helpful. Working in this way hints at the infinity of the combinations of approaches to affecting accountability mechanisms; an inability to list all possible blends of approaches to affect accountability mechanisms is not a source of embarrassment, but a recognised way of grappling with emerging domains of knowledge (Eco, 2009).

Capability Approach

As discussed in the introduction the capability approach informs a consistent goal of expanding freedom and provides an underlying ethical framework identifiable in each study. Affecting accountability mechanisms can occur without using an ethical frame or one can be selected that marginalises some to the benefit of others, especially the evaluator. Without an ethical framework the tools presented in this thesis could be applied to further evaluators' own interests. Although self-promotion is an understandable aim, unscrupulous evaluation will be rent-seeking and contribute to power imbalances, which undermines the intent of international development. There is consistent evidence that when international development neglects ethics, harm is caused, sometimes in a manner that outweighs potential good (Crush, 1995; Escobar, 2011; Gasper, 2004; Sen, 1992; Sen, 1999).

In affecting accountability mechanisms, the capability approach offers an ethical compass to apply to working in all four concepts. The capability approach defines Development as the removal of the ills of unfreedom, such as poverty and tyranny, and the expansion of the capability to undertake valued doings and beings, for example, literacy and political participation. The capability approach in pursuing freedom offers an ethical orientation focused on participation and inclusion, while engaging in the diverse values encountered in working with international development. Gasper (2002, p. 176) has summarised this position neatly: "The normative priority to capability can be read as a policy rule to promote people's capability and then 'let them make their own mistakes', rather than of achieved well-being or quality of life, though it might contribute to them." The ethical concerns of the capability approach permeate the synthesis such that it is a consistent reference point in how evaluators seek to affect an accountability mechanism.

In the synthesis two particular aspects of the ethics of the capability approach consistently surface as important: the focus on capabilities as an evaluative frame for understanding the success of development; and the importance of taking a broad informational base to support people's agency. Chronologically the first study by Porter (2009) is guided by an explicit ethical

framework based on the capability approach. In the later four studies the focus on expanding the capability of people, based on a broad information base, remains, underpinned by proximate issues such as human rights and responsiveness. The ethics of the capability approach has been subject to extensive scholarly debate (Robeyns, 2005). The synthesis presented here focuses on the ethics prioritised by the capability approach and on broadening sources of information because of the way they arise in the meta-ethnography, rather than engaging in the full range of debates about the approach.

The evaluative frame of the capability approach prioritises the attainment of freedoms (Sen, 1999). The prioritisation of freedom is consistent with international development agreements ratified by all nations (Annan and Secretary-General, 2005; UN General Assembly, 2015). When evaluating freedom in the capability approach there is interest in both the "*primary end* and…the *principal means* of development" (Sen, 1999, p. 36 with original italics). In Soviet Russia ends could be achieved, as a person was able to live in reasonable housing and their children have access to schooling, but had their means curtailed by not being able to move jobs within the labour market (Porter, 2009). Considering freedom in terms of both means and ends is first an ethical issue, then a technical issue. In working with the capability approach to affect accountability mechanisms considers both means and ends and requires evaluators to ask themselves questions that focus not only on whether objectives were achieved, but also on whether they were the right objectives and whether the right values were being pursued, otherwise known as triple-loop learning (Picciotto, 2018).

An evaluation drawing on the capability approach requires a broad base of information to evaluate both the *means* and *ends* of a development intervention (Sen, 1992; Sen, 1999; Sen, 2007). The approach of an evaluator should not reduce freedom by putting strain on participants, offending cultural norms in collecting data, but should approach power differentials in a manner that enhances the voices of marginalised people. Accessing broad sources of information through inclusive and participatory techniques means that the evaluator seeks to engage with alternative perspectives on values in both strategic and tactical processes of affecting accountability mechanisms. Taking a broad information base means that the evaluator highlights limitations in information and maintains practices that enable reasoning and deliberation of alternative values. Practically, an evaluator does not have the scope to endlessly expand the information analysed, but a stance can be taken where important information is not closed out prematurely.

Expanding the information base and examining the consistency of means and ends of an intervention supports the ability of an evaluator to examine invisible and hidden power, especially in evaluation processes. Broadening the information does not necessarily mean using new information, but can mean looking at existing information from a new angle. Drawing from Lukes (2005), existing information can help to generate counter-factual scenarios to help identify how invisible power affects the authorising environment and active exercise of power. Porter (2017), for example identifies the ability of evaluation to put an issue, such as human rights, on the policy agenda, by providing a counter-factual scenario. Taking a means-ends perspective can support identifying where the processes are unlikely to generate the official

stated ends. Porter (2009) in accordance with self-reflective approaches to evaluation (as discussed in Johnson and Rasulova, 2017) highlights that inclusive processes help to identify discrepancies between means and ends and provide hypotheses for evaluators navigating the authorising environment to identify where hidden or invisible power is at work.

The ethics of the capability approach arise in both the authorising environment and within evaluation processes. In the authorising environment the ethics require identifying and working with influential agents who will seek to surface voices of marginalised people in accountability mechanisms. In the evaluation process the evaluator seeks to shape demand, to design processes that are inclusive in their orientation, and to attend to alternate views of success in analysis. The following discussion of each approach surfaces and provides specific examples of the importance of the ethics and informational orientation of the capability approach.

Demand

The structure of demand determines how evaluation systems operate. Two important elements in understanding demand are: the policy instruments that the determine interactions in the evaluation system, and the extent to which demand for information is endogenous. When demand for information endogenously arises, agents are more likely to use information to undertake change (Bemelmans-Videc et al., 2003; Boyle and Lemarie, 1999; Chelimsky, 2006; Fukuda-Parr et al., 2002; Lopes and Theisohn, 2003; Lopez-Acevedo et al., 2012; Mackay, 2007; Pollitt et al., 2009).

Porter and Goldman (2013), provide a diagnostic toolset to analyse demand for evaluation, building on six case studies of African monitoring and evaluation systems that can be applied to assist strategic interactions when working in the authorising environment (Porter et al., 2012). The toolset analyses three areas of policy instruments: institutional design; monitoring systems; and evaluation systems. In institutional design the toolset analyses mandates, budget, and institutional location. In the monitoring and evaluation systems space, the toolset reviews coverage, frequency of reporting, plans, standards, guidelines and training. Analysing systems in this way helps to delineate the visible policy instruments that have been put in place to maintain demand for information.

Understanding the policy instruments supports an evaluator in identifying the areas of pertinence for the authorising environment (Porter and Goldman, 2013). For example, in reviewing six government monitoring and evaluation systems in Africa, Porter and Goldman (2013) note that the policy instruments primarily set-up a demand for routine monitoring information. Porter and Goldman (2013) also argue that the information gathered is inadequate in its results orientation and often does not help institutions to be responsive to the needs of the most marginalised.

Analysis of the structure of demand through the policy instruments also helps to identify the origins and structure of demand – whether it is endogenous or exogenous to the system. Equipped with this understanding, the evaluator can target potential allies in the authorising environment who require knowledge in a form that is supportive of their goal. Porter and

Goldman (2013), for example, highlight that parliaments can help to develop monitoring and evaluation information as they provide an opportunity for increased use of the findings for accountability.

Understanding the structuring of demand and explicit demand of policy instruments helps to identify consistencies and inconsistencies in the exercise of power. When accountability mechanisms make decisions and hold agents to account, insight can be gained into the structures of power. Evaluation undertaken for an exogenous donor demand, might be a cover for rent-seeking and thereby help to reveal a hidden exercise of power (Porter, 2017; Porter and Goldman, 2013). Likewise when an accountability mechanism diverges from defined policy instruments this helps to reveal contested power, whether from a hidden or invisible source (Lukes, 2005; Porter, 2013). Understanding the degree of consistency in demand can enable the evaluator to identify the extent to which demand can be stretched in the evaluation process or whether new opportunities may exist for adapted goals to be introduced in the authorising environment, as identified in Porter (2017).

Understanding information demands from the perspective of policy instruments helps undertake evaluation processes. An overall understanding of policy instruments orientates the evaluator in responding to constraints in demand and less directly in undertaking a shared journey on evaluation quality. By understanding policy intent and capacity gaps, evaluators are able to make more targeted choices during evaluations and identify who has the potential exercises power in different settings. For example, by identifying parliaments as potential users of evaluation and sources of contestation that can be engaged during evaluation processes to develop evaluative perspectives.

Evaluation systems in the cases highlighted by Porter and Goldman (2013) operate in contexts where short-term goals and the desire for control by the executive branch of government predominate, which produces monitoring and evaluation that can obfuscate performance; an issue not unusual in the operation of evaluation systems (Ryan, 2002; Ryan, 2004, 2005). The tools to analyse policy instruments applied by Porter and Goldman (2013) do help to define the structure of demand. The tools do not, however, help identify specifics around potential influential allies. The principal-agent approach needs to be layered on an analysis of demand and supply to help define the terrain of politics that surrounds the evaluation system.

Political Economy

Porter (2017) develops a toolset for political economy analysis of the context, in order to identify specific entry-points to enhance evaluation practice, especially in the authorising environment based on experience from five case studies of African Evaluation systems (Adams et al., 2013; Alemu and Latib, 2013; Kumwenda and Latib, 2013; Porter and Gasana, 2013; Porter and Mulenga, 2013). A political economy analysis assists the evaluator in understanding the predominant logic through which power is exercised in a state. The analysis of the political logic is then applied to develop a framework to unpack how agents (principals, intermediaries and agents) interact in the commissioning, managing and conduct of evaluation. Using the analysis, potential allies can be identified based upon their political position. A demand

assessment augments the political economy analysis by assessing the policy instruments and the extent to which endogenous demand is present within the evaluation system. An example helps to illustrate how principal-agent and demand analysis come together.

Porter (2017) presents a framework where intermediaries are entrusted to act on behalf of principals by managing evaluation implementation, and liaising with commissioners of the evaluation process. Intermediaries are not passive agents in the process, they are in positions where they "develop knowledge of institutions and policy that allows them to influence demand and evaluation implementation, potentially, in accordance with their own agenda" (Porter, 2017, p. 24). The intermediaries can work both strategically in the authorising environment or tactically in shaping evaluation processes. chapter highlights that intermediaries will have more or less power in different political economy contexts. In states with centralised decision-making like Rwanda, the intermediaries in the Ministry of Finance will have a large degree of power to set agendas across the state (Porter and Gasana, 2013). In countries like Zambia with a more dispersed power network, intermediaries will need to liaise with other groups in the conduct of evaluation. For example, in Zambia the Ministry of Finance sourced funding for the evaluation of its national development plan from donors (Porter and Mulenga, 2013).

Evaluators who understand the distribution of power between principals and intermediaries can better identify how the agendas of intermediaries potentially affects accountability mechanisms. Porter (2017, p. 33) describes how civil society organisations have become suppliers of evaluations that are legitimate within the power networks even in countries that are more authoritarian. In Ethiopia, for example, "the Poverty Action Network of Ethiopia has conducted evaluations of initiatives to feed into formal processes."

Principal-agent conceptions do provide an approximation of how power is exercised, which can be further developed in navigating the authorising environment or in the process of evaluation. Porter (2017, pp. 19-20) highlights how principal-agent conceptions applied to evaluation processes help to analyse visible decision making and hidden power to put an issue on the agenda to contest decisions. It shows how a view on invisible power can be developed by studying the unconscious exercise of power (Lukes, 2005). In evaluation processes or working with authorising environments, evaluators can identify and potentially help make an agent conscious of how a decision perversely effects marginalized people or benefits those in power (Porter, 2011). Where this occurs, evaluators have the opportunity to expand the focus of demand or highlight where the agenda of those with power has perverse consequences.

The ethic of the capability approach can be overlaid with a political economy analysis to help identify the types of freedom that states are open to expanding, as well as agents who have legitimate voice in specific contexts. For example, in Zambia the Jesuit Centre for Theological Reflection - a well-respected evaluation, research, education and advocacy organisation - produces a basic needs analysis based on the price of a basket of goods. The prominent position of Christianity in the Zambian constitution and in society, affords the Jesuit Centre for Theological Reflection and their evaluations a role in policy discussions. A political economy

analysis also helps to identify entry-points on both the demand and supply sides with similar ethical outlook and legitimacy in the political economy.

Helping

Agents are unlikely to invest in developing relationships with an evaluator unless they perceive something as helpful to their own agenda and their potential exercise of power. The approach of Helping (Schein, 2009) provides a basis to establish relationships, which was developed from experience in business consultancy in both the authorising environment and evaluation process. The approach argues that helpful partnerships form on the basis of equitable exchange. An evaluation process is a form of partnership. In this partnership the evaluator delivers evaluations and affects the accountability mechanism and the agent receives suitable evidence linked to a shared goal. The success of the partnership depends on how parties choose to play roles. In working in the authorising environment and in conducting evaluation processes the implication "for would-be helpers is to become conscious of social economics and the social theatre that we all live in, to think clearly about the helper role" (Schein, 2009, p. 29). Helping directs us to think about and assess the currency and values to be managed to make the engagements with agents fair and equitable (Schein, 2009).

Porter (2011) applies the lens of Helping to explain the successes and challenges of developing a monitoring system in an orphans and vulnerable children's development project in South Africa. The case analyses how an evaluator needs to consider multiple authorising environments in developing a monitoring system for the project. The authorising environment in this case consists of donors, an international non-governmental organisation (INGO), and community-based organisations' management and governance structures. Through the lens of this meta-ethnography, the case illustrates how an analysis of demand in an authorising environment helped to define approaches to expand the focus of demand and develop a shared journey on evaluation quality.

The main tool Porter (2011) discusses to develop relationships applies four forms of enquiry: pure; diagnostic; confrontational; and process-orientated. To undertake effective Helping you need to start with the first form of enquiry, pure. In pure enquiry the helper seeks to develop trust and rapport, which involves understanding what different agents are seeking to achieve and how they are currently working. Having developed trust and rapport, the helper moves to also practice diagnostic enquiry. Here the helper seeks to enable agents to peel back the layers and understand motivations, feelings and outcomes of their work. The diagnostic enquiry enables the helper to work on hunches on how to progress the agent towards their goal. Having set up the relationship and an understanding of the context, agents are confronted to make changes to their processes. In addition, the helper continues to check on the process to be assured that the process of Helping moves in the right direction. Each of these forms of enquiry are further defined in Table 6.

Table 6: Forms of Enquiry

Inquiry	Purpose	Example Questions
Pure Inquiry	• build confidence and status of the client; • develop context to reveal anxiety, feelings and information; • diagnose issues and plan for action (Schein, 2009: 69-70).	Tell me more…. When did this last happen? Can you give me some examples? (Schein, 2009: 70)
Diagnostic inquiry	Influence clients' mental processes by focusing on issues other than the ones the client chose to report, in terms of: • feelings and reactions; • causes and motives; • actions taken or completed; • systemic questions (Schein, 2009: 72-75).	How did you feel about that? - Feeling and reaction. For what reasons do you think you are having this problem? – Cause and motive. What have you tried to do so far? – Action taken. How will your colleagues react? – Systemic question (Schein, 2009: 75)
Confrontational inquiry	The helper now starts to articulate their analysis by: • making suggestions; • offering options (Schein, 2009: 75).	Did that make you angry? Could you do the following? (Schein, 2009: 76)
Process-orientated inquiry	Focusing on the clients and helper's interaction to make the client conscious of the helper's influence. This can be combined with the other forms of inquiry (E. Schein, 2009: 77)	Are we getting anywhere? Are my questions helping you? (Schein, 2009: 77)

Source: Porter (2011, p. 157)

Working through the Helping approach assists with conscientizing agents to how they operate within structures of power or have the capability to exercise power. This is because the approach aims to create a relationship that enables reflection and critical reasoning on how processes have unfolded and converge or diverge with desired ends. Porter (2011), for example, demonstrates how the process of helping enabled care organisations to identify where they were not using their capabilities to efficiently address the needs of patients. This has particular relevance in the authorising environment when establishing the legitimacy of a goal or in an evaluation process when seeking to expand the focus of demand.

The ethic of the capability approach applies to Helping by serving as an analytical tool to question the parameters of the activity. In Porter (2011) in undertaking a Helping process an orientation to expanding freedom meant that participation was broadened during the process.

In order to be helpful, the monitoring system needed to broaden the information base to provide accurate details on the processes, quantity and quality of care being delivered. In keeping with the means being consonant with the ends, the development of the monitoring system involved the staff of the community organisations. Second, the ethic of the capability approach also serves as an analytical tool to question the parameters of the case. For example, within Porter (2011) no mention is made in the case of whether the information from the monitoring system was fed back to communities or if children's lives improved because of the project.

The Helping approach to affect accountability mechanisms responds to demands for evidence and helps to shape them through supporting reflection on systems of power. An evaluator also needs to have other tools to shape evidence demands in the authorising environment and during evaluation processes. In particular the evaluator needs to be able to develop and present options on the kinds of questions the evaluation can respond to in working with commissioners of evaluation.

Joint Production

Evaluators need tools to bring together diverse agents in the authorising environments and evaluation processes aimed at influential evidence generation. Demands for evidence are diverse and different agents may not see how their agendas align. Joint production[16] serves as a tool that helps evaluators affect evaluation processes by providing tests that improve problem solving based on shared evidence.

Joint production arises as a tool within Porter (2013) to augment a rights-based approach to development. Joint production generates a shared information base by directing evidence generation to meet three tests – namely, that the evidence: needs to answer important questions; through legitimate processes; and using credible techniques (Packard Foundation, 2010a). These three tests are further detailed below.

First, evidence needs to be salient by addressing important issues. Salience can be achieved by seeking to understand the important issues of agents: What evidence will be useful to help agents meet their goals? The toolset of Helping described previously, in addition to techniques developed in theory-based evaluation, assists in identifying salient issues (Funnell, 2000; Funnell and Rogers, 2011). Undertaking pure and diagnostic enquiry assists the evaluator in understanding the issues that agents seek to address. Likewise, unpacking the theory of change helps people to be explicit about why and how they think change occurs, providing an entry-point into areas where there is uncertainty (Funnell, 2000; Funnell and Rogers, 2011). In both the authorising environment and evaluation process, evaluators would connect with agents to

[16] Joint production is also known as co-creation and co-production. In keeping with Porter Porter, S. (2013). A Change of road for a rights-based approach? A case study of mobile technology in KwaZulu Natal, South Africa. In P. Gready and W. Vandenhole. (Eds.), *Human rights and development in the new millennium: Towards a theory of change*. Routledge. the term joint production is used in this paper.

identify issues of salience and understand how these connect to their own goals to frame evaluation systems and processes.

Second, the process of evaluation needs to be seen as legitimate by agents. Legitimacy reflects the political dimension of evaluation and is founded on the credentials of an evaluators' history, expertise and process. An evaluator needs a history that demonstrates to stakeholders that they understand their political economy and cultural context. In Rwanda, for example, the Institute for Research and Dialogue for Peace (IRDP) think tank has the ability to challenge government in ways others cannot because of their history of working on issues of sustainable peace with a post-genocide understanding (Porter and Gasana, 2013). An evaluator also needs to have competencies that are perceived to be the right ones to answer the salient questions. For a Ministry of Health, for example, an evaluator with a PhD in public health and extensive experience in international development would be important criteria to make them legitimate. Equally, for an NGO the evaluator may require experience in participatory approaches rather than specific PhDs.

Third, evidence production needs to be credible. Credibility depends on the information passing appropriate methodological tests of validation. The type of methodological validation depends on the standards applied that define whether the results are seen as trustworthy. Again these can vary between different authorising environments and change over time (Porter and Hawkins, 2019). For example, an evaluation system that prioritises counter-factual evaluation designs will have different emphasis from a system that focuses mainly on the review of multi-faceted policy interventions. In an authorising environment, the evaluator would need to understand the current trends in methodological arguments, test ideas on credibility with agents and suggest new methods. In an evaluation process, the emphasis is on undertaking a shared journey on evaluation quality. To affect accountability mechanisms, evaluation processes need to identify ways to develop the capacities of agents who manage and commission quality evaluations, for example, by involving them in data collection.

Across these three tests the application of joint production in the authorising environment differs from its application in an evaluation process. An evaluation process focuses on delivering an evaluation by developing evaluation questions and applying standards of professional and methodological practice. Whereas in the authorising environment the evaluator looks to connect to agents with power to jointly produce a shared goal to strive towards over a number of years.

Jointly producing evidence provides multiple entry-points in an evaluation process for evaluators to understand and affect the invisible, hidden and visible exercise of power by accountability mechanisms. Visibly joint production has the potential to bring decision-makers into dialogue, questioning or making explicit their own demands in an explicit and publicly reasoned manner (Porter, 2013). Joint production also enables new questions to be put forward, elevating potentially hidden debates to the forefront of developing and implementing an evaluation. Joint production also enables evaluators to actively facilitate or observe how people respond when they have the opportunity to move to a peer rather than subordinate

position in a hierarchy and so better identify invisible constraints. Understanding how marginalized people, for example, do and do not take opportunities can help to identify changes in how, for example, evaluation quality is understood. If opportunities are not taken to shape evaluation processes to take into account marginalized people's perspectives, then changes in overall evaluation processes can be suggested to make these perspectives more prominent, as was attempted in Porter and de Wet (2009).

The ethic of the capability approach informs the boundaries and emphasises joint production. In Porter (2013), for example, the application of joint production is proposed in a context that centralises the furthering of rights-based development, a specific ethical frame that complements the capability approach (Sen, 2005). Enshrined in international law and employed by UN agencies rights-based development focuses on government becoming more responsive to citizens through recognising their legal claims (Gready and Ensor, 2005). Improving rights-based development through applying joint production to an authorising environment, emphasises enhancing freedom. The enhancement of freedom occurs through joint production broadening the informational base and connecting different agents – citizens and government – to enhance dialogue and responsiveness of government.

6.8 Summary of the Synthesis

Through this line-of-sight synthesis we can see the concepts and approaches work together to layer insights and suggest tools and ways of working in both the authorising environment and evaluation process. Each of the four concepts plays a specific role in affecting accountability mechanisms. The different approaches presented here also interact, help analyse and potentially influence the exercise of power by revealing invisible, hidden and visible exercises in power whether: defining counter-factual alternatives; identifying disjuncture's between intent and action; highlight the unconscious exercise of power; or understanding how opportunities for contestation happen or are missed. In activating two or more concepts simultaneously, as they are in the studies, it can be inferred that they are more likely to affect accountability mechanisms. In turn this will influence the use of evaluation products. Table 7 below summarises the four concepts developed through the meta-ethnography and their where, why, what and how.

The approaches to working in each of the concepts need to be blended to ensure successful implementation. Working to the capability approach provides ethical coherence across the concepts. By understanding the structure of demand, the institutional configuration can be better responded to. Through applying a principal-agent lens, the politics of information and the parameters of accepted evidence are better understood. Helping provides a frame for interaction with agents, while joint production frames these interactions around the development and influence of evidence. No single approach across the concepts would affect the authorising environment; the synthesis shows us that these approaches deployed as a toolset enables an evaluator to work ethically, define entry-points, and further develop the relationships with agents with power.

Table 7: Line-of-Sight Summary

Concept	Why?	What changes?	How? (examples)_
Authorising environment			
Navigating the authorising environment based on a goal	Develop a shared vision of change within a network of agents on the chosen goal	Evaluators focus on how an authorising environment's network operates and where there are demands to improve information flows and leverage the power of other agents.	Evaluators require competencies to understand the political economy of the demand for evaluation and to work between agents in a helpful manner so they can connect on shared goals and ethics.
Evaluation Processes			
Expand the focus of demand	Achieving the chosen goal means adapting the structure of demand in given constraints	Understand constraints and seek to change demand by: • Giving voice to the target group • Nurture incipient demand for evidence	Evaluators broaden the relevant information base to shift focus towards performance, rather than narrowly prioritising control.
Access and respect the agenda of agents	Interests need to be understood to provide influential findings	Agents' agendas are understood by: • Inclusive discussion of agendas • Connecting the evaluation to the roles of principals, intermediaries and agents.	Evaluators with competencies in action learning, political economy and participation can better understand agents' diverse perspectives and desires, and can frame arguments that resonate with these.
Undertake a shared journey on evaluation quality	Evaluation capacity and quality improves agent's interaction with evaluation	Developing capacities in individual and collective evaluative thinking	Evaluators develop standards to enable agents to understand appropriate forms of evaluation. Through this emphasis on capacity, evaluators and agents undertake co-design, training, reflection and networking together.

7. Conclusion

People with power often publicly demand accountability. A search of many politicians reveals a quote where they extol or demand accountability, regardless of their political persuasion (Al-Saud, 2020; BBC, 2018; CNN, 2019; Kwibuka, 2018). One place where these demands are played out are accountability mechanisms, where agents meet to exercise and constrain power. For evaluators interested in working with accountability mechanisms it is a challenge to shape the way the mechanism operates to support the achievement of a goal – be it freedom, responsiveness, evidence use, or transformation. Overcoming this challenge requires relationships with those with power and the surfacing invisible power in the form of contestation, unconscious exercises, counter-factual scenarios and disjuncture with stated goals of poverty reduction.

7.1 The Journey so Far

This meta-ethnography has sought to develop a middle-range theory of how evaluation affects accountability. All seven phases of a meta-ethnography have been worked through. A literature review identified a research gap around how evaluation affects accountability mechanisms (phase 1). Areas of relevance were specified by defining objectives and questions (phase 2). Five articles and chapters that I authored or co-authored over a period of ten-years were selected to serve as the basis for the meta-ethnography. The substance of the meta-ethnography was undertaken with papers being read, re-read, concepts juxtaposed, interpretations and contrasts made, and a synthesis developed (phases 3-6). The synthesis was expressed by describing in where, why, what and how four defined concepts operate. Together these concepts define a middle-range theory that can guide evaluators in affecting accountability mechanisms.

The study has found that there are different ways for evaluators to work with accountability mechanisms. Evaluators can produce a report that accountability mechanisms use as an instrument in their deliberations. This is the subject of much of the current evaluation literature and focuses attention more on the operational implementation of evaluation. Evaluators can work tactically to support changes in the use of a report. Alternatively, evaluators can combine efforts tactically in evaluation processes and strategically in the authorising environment to develop how accountability mechanisms define their criteria and standards, assess progress, and synthesise information in making consequential decisions.

By working substantively with accountability mechanisms, evaluators adapt how they operate and how the power to dominate others is shaped and exercised. A middle range theory has been proposed in this thesis that anticipates evaluation affects accountability by working across four concepts: (i) strategically navigating the authorising environment based on a goal; and tactically in evaluation processes (ii) expanding the focus of demand, (iii) accessing and respecting the agenda of agents and (iv) undertaking a shared journey on evaluation quality. When these four concepts are activated, accountability mechanisms adapt the processes used to make consequential decisions.

In working across the one strategic and three tactical concepts, evaluators are engaging in the space of politics, where power is exercised and constrained. Working in the space of politics requires different skills from working in a technically proficient manner. As Chelimsky (2006, p. 43) states in reference to working on politically charged issues, “often the context is more like the fog of war surrounding a pitched battle.” In working with the four concepts an evaluator will require competencies in different approaches and theories. Evaluators need to competently understand demand and the political economy of the context, practice Helping in change processes, and undertake joint production. The evaluator also requires an ethical approach when engaging accountability mechanisms to keep them focused on coherence of the means with the ends they seek to attain. In this study the guiding ethic was Sen’s (1999) capability approach, which prioritises freedom, that emerged as an important frame in affecting accountability mechanisms. Across these approaches and concepts Lukes’ (2005) three-dimensional view, which prioritises the exploration of power as a form of domination helps to identify how accountability mechanisms operate and ways they can be affected.

7.2 Contribution and Application

This meta-ethnography contributes an articulation of an empirically grounded middle-range theory that supports evaluators, agents working with accountability mechanisms to affect their functioning. Further academics studying evaluation and accountability can further test the middle range theory and further explore the use of meta-ethnography.

The gap closed by this theory also provides an opportunity to adapt the way the discipline of evaluation views and approaches accountability. Evaluators often discuss the difference between the learning and accountability purposes of evaluation (e.g. Rossi et al, 2018; Christie and Alkin, 2013). This middle range theory extends the conversation by showing accountability mechanisms can take a variety of perspectives that can reinforce reflection or sanctions depending on the goal being sought. Recognising the substantive role that evaluators can play in accountability mechanisms especially in identifying and surfacing invisible and hidden power also expands discussions of evaluation beyond its role as an instrument in accountability mechanisms. In addition, evaluation professionals can further explore how the logic of evaluation affects other practices, be those management or decision-making processes. Elements of this expansion are already occurring in the discussion on evaluative thinking (Buckley *et al.*, 2015).

Agents working with accountability mechanisms to further evidence-based practice can also explore the application of the middle-range theory presented in this work. Intermediaries who work between principals and agents (discussed in section 6.7) can work between evaluators and those with power in accountability mechanisms to innovate with both their practices. For example, intermediaries applying the theory developed here would expand the focus of demand of evaluators and agents in accountability mechanisms and shape evaluation quality towards a common goal.

The study provides an example of the development of a middle-range theory using a meta-ethnography approach that can be of interest to academics studying evaluation and accountability. As highlighted in the introduction, middle-range theory has garnered interest in those funding evaluations. The presentation here provides an example for how synthesis can be conducted on existing research using a meta-ethnography approach to generate a middle-range theory. The discussion here also shows that the connection between power, accountability and evaluation runs deep. All three depend to some extent on each other occurring in dispositional, relational and evaluative forms. These traits are not unique to power, accountability and evaluation. These three facets can provide a useful framework for studying the connection and interplay in other policy processes, such as agenda setting and design.

7.3 Areas for Further Investigation

The main limitation of this study is that it does not directly address the effects of evaluation on accountability mechanisms, such as use or the extent to which processes change or whether shifts in power hold. There are three areas of follow-up that emerge from this limitation.

First, to understand the effect of evaluation on accountability mechanisms requires additional investigation. This study can only infer that the use of evaluation will be improved through evaluators purposefully making a closer connection and building relationships between the evaluation and the substance of decision-making processes in the accountability mechanism. Further studies could explore the utilisation of evaluation and other changes in the potential exercise of power, focused on how evaluators worked with accountability mechanisms.

Second, investing in applying meta-ethnography and other synthesis approaches to develop evaluation theory will return further insights. As highlighted in section 4.2, there are limitations in empirical research in the discipline of evaluation (Blake and Ottoson, 2009; Miller, 2010; Ottoson, 2009; Patton, 2008; Shadish et al., 1991; Weiss, 1998). The application of meta-ethnography in this study has shown that it is a useful approach to understand practitioner experience, providing an alternative to recounting experience or claiming connections between concepts on the basis of they often occur together. The meta-ethnographic synthesis enables the emergence of nuance and contrasts to refine and generate new concepts. Many evaluators have portfolios of publications that, if interpreted in a similar manner, could provide new insights on practice. Elaboration of evaluation approaches can be undertaken by repurposing existing evidence, and also using similar techniques such as meta-analysis and realist synthesis. Transparent synthesis of experience exposes the working of concepts, providing a way in which other practitioners can rearrange and connect concepts to expand and tighten the explanation of how evaluation works.

Finally, the middle-range theory proposed in this study could be further tested and refined. Application of the proposed middle-range theory either to an existing case or in building an evaluation system will confirm, add to or disprove its application in different contexts. The case studies where power and evidence based-policy have been discussed provide an existing body of knowledge that could be re-examined in light of this middle-range theory (for example, Goldman

and Pabari, 2020; Guijt, 2015; Hayman et al., 2016; Hopson and Cram, 2018), or else the new studies could attempt an application. Re-examining or attempting new studies could take a more general view of power moving beyond the view of power as domination worked through here. The education setting in the US where existing research and differential experience in connecting evaluation to accountability mechanisms, could be a useful case to develop the middle-range theory (Ryan, 2002; Ryan, 2004, 2005; Ryan et al., 1996).

7.4 Final Thoughts on Ethics

This study notes the importance of the interaction of power and evaluation. Consequently, the sometimes lofty goals of evaluation, be they social justice or transformation (Mertens, 2009; Patton, 2019), are often sought in contexts where pecuniary desires, narrow elite interests, and entitlement based on class, race or gender hold sway. I have seen (as described under section 2.1) these countervailing forces produce hypocrisy traps where competing goals drive behaviours simultaneously in different directions (Runciman, 2018; Weaver, 2008). An evaluator, for example, can simultaneously affect an accountability mechanism to apply a rights-based perspective, while distorting perspectives to gain market advantage over a rival.

Ethics has been a theme throughout this study based upon Sen's capability approach and other writings (Sen, 1999; Sen, 2007; Sen, 2008). In interacting in spaces of power, such as accountability mechanisms, it becomes increasingly important to guard against deleterious spirals in our ethics. Evaluators may be looking to affect power and agents with power will also seek to affect evaluators. Our ethics are important for the maturing of professional evaluation as "no one can ever float something full of holes" (Murakami, 2019, p. 324). An ethic guards our behaviour and reminds us of what is proper, preventing rent-seeking and promoting well-intentioned efforts. Neglecting ethics causes harm (Crush, 1995; Escobar, 2011; Gasper, 2004; Sen, 1992; Sen, 1999). In undertaking evaluation practice and working with power across visible, hidden and invisible spaces during these tumultuous times of global pandemic, it remains important to pause, question and strive to enhance our ethics.

Bibliography

Ackerman, J. (2004). Co-Governance for Accountability: Beyond "Exit" and "Voice". *World Development*, *32*(3), 447-463. https://doi.org/https://doi.org/10.1016/j.worlddev.2003.06.015

Adams, S., Amoatey, C., Taabazuing, J., & Feinstein, O. (2013). *Study on the Demand for and Supply of Evaluation in Ghana*. Johannesburg: Graduate School of Public and Development Management, University of the Witwatersrand

Ake, C. (2001). *Democracy and development in Africa*. Washington: Brookings institution press.

Alemu, G., & Latib, S. (2013). *Study on the Demand for an Supply of Evaluation in Ethiopia*. Johannesburg: Graduate School of Public and Development Management, University of the Witwatersrand. Available from: https://www.theclearinitiative.org/resources/study-on-demand-for-and-supply-of-evaluation-ethiopia, accessed on: 14 April 2020

Alkin, M. C., & Christie, C. (2004). 'An Evaluation Theory Tree'. In M. C. Alkin (Ed.), *Evaluation Roots: Tracing Theorists Views and Influences* (pp. 12-65). Thousand Oaks: SAGE

Alkire, S. (2005). *Valuing freedoms: Sen's capability approach and poverty reduction*. Oxford: Oxford University Press

Al-Saud, M.B.S. (2020). *Message From HRH Prince Mohammed Bin Salman Bin Abdulaziz Al-Saud*. Available from: https://vision2030.gov.sa/en/vision/crown-message, accessed on: 13 July 2020

Annan, K. A. (2005). *In Larger Freedom: Towards Development, Security and Human Rights for All: Report of the Secretary-General: United Nations*, 21 March 2005, A/59/2005. Available from: https://undocs.org/A/59/2005, accessed on: 13 July 2020

Atkins, S., Lewin, S., Smith, H., Engel, M., Fretheim, A., & Volmink, J. (2008). 'Conducting a meta-ethnography of qualitative literature: Lessons learnt'. *BMC Medical Research Methodology*, *8*(1), 21. Available from: https://doi.org/10.1186/1471-2288-8-21

Barnett, M., & Duvall, R. (2005). Power in International Politics. *International Organization*, *59*(1), 39-75. https://doi.org/10.1017/S0020818305050010

Barnett-Page, E., & Thomas, J. (2009). 'Methods for the synthesis of qualitative research: a critical review'. *BMC Medical Research Methodology*, *9*(1), 59. Available from: https://doi:10.1186/1471-2288-9-59

Batliwala, S. (2019). *All about Power: Understanding Social Power and Power Structures.* New Dehli: Crea.

BBC News (2018). 'Labour MPs 'now accountable to members''. Available from: https://www.bbc.com/news/av/uk-politics-44143082/labour-mps-now-accountable-to-members, accessed on: 24 May 2020

Behn, R. D. (2001). *Rethinking democratic accountability*. Washington, DC: Brookings Inst Press.

Behn, R. D. (2003). 'Why Measure Performance? Different Purposes Require Different Measures'. *Public Administration Review, 63*(5), 586-606.

Behn, R. D. (2014). 'PerformanceStat: Accounting for Targeted Performance'. In M. Bovens, R.E. Goodin, & T. Schillemans (Eds.), *The Oxford Handbook of Public Accountability* (pp 456-471). Oxford: Oxford University Press

Bell, S., & Aggleton, P. (2016). *Monitoring and evaluation in health and social development: interpretive and ethnographic perspectives*. London: Routledge. https://doi.org/10.4324/9781315730592

Bemelmans-Videc, M. L., Lonsdale, J., & Perrin, B. (2007). *Making accountability work: Dilemmas for evaluation and for audit (Vol. 14)*. New Brunswick, NJ: Transaction Publishers

Bemelmans-Videc, M. L., Rist, R. C., & Vedung, E. (2003). *Carrots, sticks, and sermons: Policy instruments and their evaluation*. New Brunswick, NJ: Transaction Publishers

Benington, J., & Moore, M. H. (2011). 'Public Value in Complex and Changing Times'. In J. Benington & M. H. Moore (Eds.), *Public Value: Theory and Practice* (pp. 1-30). London: Palgrave Macmillan.

Benjamin, L. M. (2008). 'Evaluator's role in accountability relationships: Measurement technician, capacity builder or risk manager?' *Evaluation, 14*(3), 323-343.

Berger, B. K. (2005). Power over, power with, and power to relations: Critical reflections on public relations, the dominant coalition, and activism. *Journal of Public Relations Research*, *17*(1), 5-28.

Blake, S. C., & Ottoson, J. M. (2009). 'Knowledge utilization: Implications for evaluation'. *New Directions for Evaluation, 2009(*124), 21-34.

Bovens, M. (2007). 'Analysing and Assessing Accountability: A Conceptual Framework'. *European Law Journal, 13*(4), 447-468. Available from: https://doi:10.1111/j.1468-0386.2007.00378.x

Bovens, M. (2010). Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism. *West European Politics*, *33*(5), 946-967. https://doi.org/10.1080/01402382.2010.486119

Bovens, M. (2014). 'Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism'. In D. Curtin, P. Mair & Y. Papadopoulos (Eds.), *Accountability and European Governance* (pp. 28-49). Abingdon: Routledge.

Bovens, M., Goodin, R. E., & Schillemans, T. (2014a). 'Public Accountability'. In M. Bovens, R. E. Goodin, & T. Schillemans (Eds.), *The Oxford Handbook of Public Accountability* (pp. 1-23). Oxford: Oxford University Press.

Bovens, M., Goodin, R. E., & Schillemans, T. (2014b). *The Oxford Handbook of Public Accountability* Oxford: Oxford University Press.

Boyle, R., Lemaire, D., & Rist, R. C. (1999). 'Introduction: Building evaluation capacity'. In R. Boyle & D. Lemaire (Eds.), *Building Effective Evaluation Capacity: Lessons from Practice* (pp. 1-22). New Brunswick, NJ: Transaction Publishers.

Boyle, R., & Lemaire, D. (1999). *Building Effective Evaluation Capacity: Lessons from Practice*. New Brunswick, NJ: Transaction Publishers.

Britten, N., Campbell, R., Pope, C., Donovan, J., Myfanwy, M., & Roisin, P. (2002). 'Using meta ethnography to synthesise qualitative research: a worked example'. *Journal of Health Services Research & Policy, 7*(4), 209-215. Available from: https://doi:10.1258/135581902320432732

Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). 'Defining and Teaching Evaluative Thinking: Insights from Research on Critical Thinking'. *American Journal of Evaluation, 36*(3), 375-388. Available from: https://doi:10.1177/1098214015581706

Campbell, R., Pound, P., Morgan, M., Daker-White, G., Britten, N., Pill, R., Yardley, L., Pope, C. & Donovan, J. (2011). 'Evaluating meta ethnography: systematic analysis and synthesis of qualitative research'. *Health Technology Assessment, 15*(43), 164. Available from: https://doi:10.3310/hta15430.

Campbell, R., Pound, P., Pope, C., Britten, N., Pill, R., Morgan, M., & Donovan, J. (2003). 'Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care'. *Social science & Medicine, 56*(4), 671-684. Available from: https://doi: 10.1016/s0277-9536(02)00064-3

Centre of Excellence for Development Impact and Learning. (2020). *Unveiling 25 Innovative CEDIL Projects: Breaking New Ground on Design and Filling Evidence Gaps*. Available from: https://cedilprogramme.org/unveiling-25-innovative-cedil-projects/, accessed Feb 6 2021

Charity Commission For England and Wales (2019) *Statement of Results of an Enquiry Oxfam Registered Charity Number 202918.* Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/807945/Statement_of_the_Results_of_an_Inquiry_Oxfam.pdf, accessed Feb 6 2021

Chelimsky, E. (2006). 'The Purpose of Evaluation in a Democratic Society'. In I. Shaw, J. C. Green, & M. M. Mark (Eds.), *The Sage Handbook of Evaluation* (pp. 33-55). Thousand Oaks: SAGE

Chen, H-T. (2005). *Practical Program Evaluation: Assessing and Improving Planning, Implementation and Effectiveness*. Thousand Oaks: SAGE

Chilisa, B., Major, T. E., Gaotlhobogwe, M., & Mokgolodi, H. (2015). 'Decolonizing and indigenizing evaluation practice in Africa: toward African relational evaluation approaches'. *Canadian Journal of Program Evaluation, 30*(3), 313-328. Available from: https://doi: 10.3138/cjpe.30.3.328

Chouinard, J. A. (2013). 'The Case for Participatory Evaluation in an Era of Accountability'. *American Journal of Evaluation, 34*(2), 237-253. Available from: https://doi:10.1177/1098214013478142

Christie, C. A., & Alkin, M. C. (2013). 'An Evaluation Theory Tree'. In M. C. Alkin (Ed.), *Evaluation Roots: A Wider Perspective of Theorists Views and Influences* (pp. 11-57). Thousand Oaks: SAGE

Clark, W. C., Mitchell, R. B., & Cash, D. W. (2006). 'Evaluating the Influence of Global Environmental Assessments'. In R. B. Mitchell, W. C. Clark, D. W. Cash, & N. Dikinson (Eds.), *Global Environmental Assessments: Information and Influence* (pp. 1). Cambridge: MIT.

Clayton, D. M. (2018). Black Lives Matter and the Civil Rights Movement: A Comparative Analysis of Two Social Movements in the United States. *Journal of Black Studies*, *49*(5), 448-480. https://doi.org/10.1177/0021934718764099

CNN. (2019). *Bipartisan senators again aim to hold Saudi Arabia accountable*. Available from: https://www.cnn.com/2019/07/10/politics/saudi-arabia-accountability-bill/index.html, accessed on: 24 May 2020

Cornwall, A. (2016). 'Women's empowerment: What works?' Journal of International Development, *28*(3), 342-359.

Cornwall, A., & Aghajanian, A. (2017). How to Find out What's Really Going On: Understanding Impact through Participatory Process Evaluation. *World Development*, *99*, 173-185. https://doi.org/https://doi.org/10.1016/j.worlddev.2017.07.010

Crack, A. M. (2013). 'INGO accountability deficits: The imperatives for further reform'. *Globalizations, 10*(2), 293-308.

Crush, J. S. (1995). *Power of Development*. Hove: Psychology Press.

Davidson, E. J. (2005). *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation*. Thousand Oaks: SAGE.

Diamond, L. J., & Morlino, L., (2005). *Assessing the Quality of Democracy*. Baltimore: Johns Hopkins University Press.

Duflo, E., & Banerjee, A. (2017). *Handbook of Field Experiments*. Amsterdam: North Holland.

Eco, U. (2009). *The Infinity of Lists: An Illustrated Essay*. New York: Rizzoli.

Eisenhardt, K. M. (1989). 'Building Theories from Case Study Research'. *The Academy of Management Review, 14*(4), 532-550. Available from: https://doi:10.2307/258557

Escobar, A. (2011). *Encountering Development: The Making and Unmaking of the Third World* (Vol. 1). Princeton, NJ: Princeton University Press.

Eyben, R., Guijt, I., Roche, C., & Shutt, C. (Eds.). (2015). *The politics of evidence and results in international development: Playing the game to change the rules?*. Rugby: Practical Action Publishing.

Feinstein, N. O. (2012). 'Evaluation as a learning tool'. *New Directions for Evaluation, 2012*(134), 103-112.

Fetterman, D. M., & Wandersman, A. (2005). *Empowerment Evaluation Principles in Practice*. New York: Guilford Press.

Fox, J. A. (2015). Social Accountability: What Does the Evidence Really Say? *World Development*, *72*, 346-361. https://doi.org/https://doi.org/10.1016/j.worlddev.2015.03.011

Fox, J. A. (2018). The Political Constrution of Accountability Keywords. *IDS Bulletin*, *49*(2).

Fournier, D. M. (1995). 'Establishing evaluative conclusions: A distinction between general and working logic'. *New Directions for Program Evaluation, 1995*(68), 15-32. Available from: https://doi.org/10.1002/ev.1017

France, E. F., Cunningham, M., Ring, N., Uny, I., Duncan, E. A. S., Jepson, R. G., . . . Noyes, J. (2019). 'Improving reporting of meta-ethnography: the eMERGe reporting guidance'. *BMC Medical Research Methodology, 19*(1), 25. Available from: https://doi:10.1186/s12874-018-0600-0

France, E. F., Ring, N., Thomas, R., Noyes, J., Maxwell, M., & Jepson, R. (2014). 'A methodological systematic review of what's wrong with meta-ethnography reporting'. *BMC Medical Research Methodology, 14*(1), 119. Available from: https://doi:10.1186/1471-2288-14-119

Fukuda-Parr, S., Lopes, C., & Malik, K. (2002). *Capacity for Development: New Solutions to Old Problems*. London: Earthscan.

Fukuyama, F. (2014). *Political order and political decay: From the industrial revolution to the globalization of democracy*. New York: Macmillan.

Funnell, S. C. (2000). 'Developing and Using a Program Theory Matrix for Program Evaluation and Performance Monitoring'. *New Directions for Evaluation, 87*, 91-101.
Funnell, S. C., & Rogers, P. J. (2011). *Purposeful Program Theory: Effective use of theories of change and logic models*. San Francisco: Jossey-Bass.

Gasper, D. (2002). 'Is Sen's capability approach an adequate basis for considering human development?' *Review of Political Economy, 14*(4), 435-461.

Gasper, D. (2004). *The Ethics of Development*. Edinburgh: Edinburgh University Press.

Gaventa, J. (1982). *Power and powerlessness: Quiescence and rebellion in an Appalachian valley*. University of Illinois Press.

Gaventa, J. (2006). Finding the spaces for change: a power analysis. *IDS bulletin*, *37*(6), 23-33.

Geertz, C. (1973). 'Thick description: Towards an Interpretive Theory of Cultures'. In C. Geertz (Ed.), *The Interpretation of Cultures* (pp. 3-31). New York: Basic Books.

Gehman, J., Glaser, V. L., Eisenhardt, K. M., Gioia, D., Langley, A., & Corley, K. G. (2018). 'Finding theory–method fit: A comparison of three qualitative approaches to theory building'. *Journal of Management Inquiry, 27*(3), 284-300.

Goldman, I., & Pabari, M. (2020). *Using Evidence in Policy and Practice: Lessons from Africa*. Taylor & Francis.

Gready, P., & Ensor, J. (2005). *Reinventing Development? Translating Rights-Based Approaches from Theory into Practice*. London: ZED Books.

Greene, J. C. (1999). 'The Inequality of Performance Measurements'. *Evaluation, 5*(2), 160-172.

Greene, J. C. (2006). 'Evaluation, Democracy, and Social Change'. In I.F. Shaw, J.C. Greene & M.M. Mark (Eds.), *The SAGE Handbook of Evaluation* (pp. 118-140). Thousand Oaks: SAGE.

Guba, E. G., & Lincoln, Y. S. (1994). 'Competing paradigms in qualitative research'. In Eds. N.K. Denzin and E.G. Guba *Handbook of qualitative research*, (pp. 105-117). Thousand Oaks, CA. Sage

Guijt, I. (2010). 'Exploding the myth of incompatibility between accountability and learning'. In J. Ubels, N-A. Acquaye-Baddoo & A. Fowler (Eds.) *Capacity Development in Practice*, (pp. 277-291). Abingdon: Routledge.

Guijt, I. (2015). Playing the rules of the game and other strategies. in eds Eyben, R., Guijt, I., Roche, C., & Shutt, C. (Eds.). *The politics of evidence and results in international development: Playing the game to change the rules?*. Rugby: Practical Action Publishing. pp. 193-209.

Hancock, D. R., & Algozzine, B. (2017). *Doing Case Study Research: A practical guide for beginning researchers.* New York: Teachers College Press.

Hawkins, P. (2010). 'Successful evaluation management: engaging mind and spirit'. *The Canadian Journal of Program Evaluation, 25*(3), 27-36.

Hayman, R., King, S., Kontinen, T., & Narayanaswamy, L. (2016). *Negotiating knowledge: Evidence and experience in development NGOs*. Practical Action Publishing Rugby.

Haugaard M. Reflections on Seven Ways of Creating Power. *European Journal of Social Theory*. 2003;6(1):87-113. doi:10.1177/1368431003006001562

Haugaard, M. (Ed.). (2002). *Power: A reader*. Manchester: Manchester University Press.

Hayne, C., & Salterio, S. E. (2014). 'Accounting and Auditing'. In M. Bovens, R.E. Goodin, & T. Schillemans (Eds.), *The Oxford Handbook of Public Accountability* (pp. 421-440). Oxford: Oxford University Press.

Hayward, C., & Lukes, S. (2008). Nobody to shoot? Power, structure, and agency: A dialogue. *Journal of Power*, *1*(1), 5-20.

Hedrick, T. E. (1988). The Interaction of Politics and Evaluation. *Evaluation Practice*, *9*(3), 5-14. https://doi.org/10.1177/109821408800900301

Hobbes, T. (1985). *Leviathan*. London: Penguin.

Holvoet, N., & Dewachter, S. (2013). 'Multiple Paths to Effective National Evaluation Societies: Evidence from 37 Low- and Middle-Income Countries'. *American Journal of Evaluation, 34*(4), 519-544. Available from: https://doi:10.1177/1098214013497083

Honig, D., & Pritchett, L. (2019). *The Limits of Accounting-Based Accountability in Education (and Far Beyond): Why More Accounting Will Rarely Solve Accountability Problems*. RISE Working Paper. Available from: https://www.cgdev.org/sites/default/files/limits-accounting-based-accountability-education-and-far-beyond-why-more-accounting-will.pdf, accessed on 14 July 2020

Hopson, R. (2003). *Overview of Multicultural and Culturally Competent Program Evaluation*. Oakland, CA: Social Policy Research Associates. Available from: http://www.calendow.org/uploadedFiles/Publications/Evaluation/Multicultural_Health_Evaluation/OverviewB, accessed on: 18 April 2020

Hopson, R., & Cram, F. (2018). *Tackling Wicked Problems in Complex Ecologies: The Role of Evaluation*. Redwood City: Stanford University Press.

House, E. R. (1980). *Evaluating with Validity*. Thousand Oaks: SAGE.

House, E. R., & Howe, K. R. (1999). *Values in Evaluation and Social Research*. Thousand Oaks: SAGE.

Jasanoff, S. (2004). *States of Knowledge: The Co-Production of Science and the Social Order.* Abingdon: Routledge.

Johnson, S. and Rasulova, S. (2017) Qualitative research and the evaluation of development impact: incorporating authenticity into the assessment of rigour, Journal of Development Effectiveness, 9:2, 263-276, DOI: 10.1080/19439342.2017.1306577

Kachur, D., Soal, S., & Van Blerk, R. (2016). 'Stretching between learning and accountability: Experiences of South African non-governmental organisations'. *African Evaluation Journal, 4*(1) Available from: https://doi:10.4102/aej.v4i1.71

King, J. A. (2007). 'Developing Evaluation Capacity through Process Use'. *New Directions for Evaluation*, *2007*(116), 45-59. Available from: https://doi:10.1002/ev.242

King, J. A., & Alkin, M. C. (2019). 'The Centrality of Use: Theories of Evaluation Use and Influence and Thoughts on the First 50 Years of Use Research'. *American Journal of Evaluation, 40*(3), 431-458. Available from: https://doi:10.1177/1098214018796328

Kirkhart, K. E. (2000). 'Reconceptualizing evaluation use: An integrated theory of influence'. *New Directions for Evaluation, 2000*(88), 5-23.

Kissinger, H. (2020). 'The coronavirus pandemic will forever alter the world order'. *Wall Street Journal* Retrieved December 24 from https://www.wsj.com/articles/the-coronavirus-pandemic-will-forever-alter-the-world-order-11585953005

Kitschelt, H., Freeze, K., Kolev, K., & Wang, Y.-T. (2020). Measuring democratic accountability: An initial report on an emerging data set. *Revista de Ciencia Política*, *29*(3), 741-773.

Kogen, L. (2018). 'What have we learned here? Questioning accountability in aid policy and practice'. *Evaluation, 24*(1), 98-112. Available from: https://doi:10.1177/1356389017750195

Kuhn, T. S. (2012). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Kumwenda, H., & Latib, S. (2013). *Study on the Demand for and Supply of Evaluation in Malawi*. In. Johannesburg: Graduate School of Public and Development Management, University of the Witwatersrand.

Kusek, J. Z., & Rist, R. C. (2004). *Ten Steps to a Results-Based Monitoring and Evaluation System: A Handbook for Development Practitioners*. Washington, DC: World Bank.

Kushner, S. (2002). 'The object of one's passion: Engagement and community in democratic evaluation'. *Evaluation Journal of Australasia, 2*(2), 16-22.

Kwibuka, E. (2018). Embrace accountability, Kagame tells leaders. *The New Times* 29 March 2018. Available from https://www.newtimes.co.rw/rwanda/embrace-accountability-kagame-tells-leaders

Lagarde M, Haines A, Palmer N. Conditional Cash Transfers for Improving Uptake of Health Interventions in Low- and Middle-Income Countries: A Systematic Review. *JAMA.* 2007;298(16):1900–1910. doi:10.1001/jama.298.16.1900

Lee, R. P., Hart, R. I., Watson, R. M., & Rapley, T. (2015). 'Qualitative synthesis in practice: some pragmatics of meta-ethnography'. *Qualitative Research, 15*(3), 334-350. Available from: https://doi:10.1177/1468794114524221

Lennie, J., & Tacchi, J. (2014). 'Bridging the divide between upward accountability and learning-based approaches to development evaluation: Strategies for an enabling environment'. *Evaluation Journal of Australasia, 14*(1), 12-23.

Lindberg, S. I. (2013). 'Mapping accountability: core concept and subtypes'. *International Review of Administrative Sciences, 79*(2), 202-226.

Lopes, C., & Theisohn, T. (2003). *Ownership, Leadership and Transformation: Can We Do Better for Capacity Development?* London: UNDP & Earthscan.

Lopez-Acevedo, G., Krause, P., & Mackay, K. (2012). *Building Better Policies: The Nuts and Bolts of Monitoring and Evaluation Systems.* Washington, DC: World Bank.

Lukes, S. (2005). *Power: A Radical View* (Second ed.). London: Springer https://doi.org/10.2307/41802305

Mackay, K. (2007). *How to Build M&E Systems to Better Support Government*. Washington, DC: World Bank.

Malpass, A., Shaw, A., Sharp, D., Walter, F., Feder, G., Ridd, M., & Kessler, D. (2009). '"Medication career" or "Moral career"? The two sides of managing antidepressants: A meta-ethnography of patients' experience of antidepressants'. *Social Science & Medicine, 68(*1), 154-168. Available from: https://doi.org/10.1016/j.socscimed.2008.09.068

Manroth, A., Hernandez, Z., Masud, H., Zakhour, J., Rebolledo, M., Mahmood, S. A., Seyedian, A., Hamad, Q., & Peixoto, T. (2014). *Strategic framework for mainstreaming citizen engagement in World Bank Group Operations: Engaging with citizens for improved results*

Mark, M. M., & Henry, G. T. (2004). 'The mechanisms and outcomes of evaluation influence'. *Evaluation, 10*(1), 35-57.

Mark, M. M., Henry, G. T., & Julnes, G. (1999). 'Toward an integrative framework for evaluation practice'. *American Journal of Evaluation, 20*(2), 177-198.

Marra, M. (2018). 'The ambiguities of performance-based governance reforms in Italy: Reviving the fortunes of evaluation and performance measurement'. *Evaluation and Program Planning, 69,* 173-182. Available from: https://doi.org/10.1016/j.evalprogplan.2017.02.006

Mayne, J., Divorski, S., & Lemaire, D. (1999). 'Locating Evaluation: Anchoring Evaluation in the Executive or Legislature, or Both or Elsewhere?' In R. Boyle & D. Lemaire (Eds.), *Building Effective Evaluation Capacity: Lessons from Practice*. New Brunswick, NJ: Transaction Publishers.

McGee, R., & Pettit, J. (2019). *Power, Empowerment and Social Change*. Taylor & Francis Group.

Mertens, D. M. (2009). *Transformative Research and Evaluation*. New York: Guilford Press.

Merton, R. K., & Merton, R. C. (1968). *Social Theory and Social Structure*. New York: Simon and Schuster.

Miller, R. L. (2010). 'Developing standards for empirical examinations of evaluation theory'. *American Journal of Evaluation, 31*(3), 390-399.

Moncrieffe, J. M. (1998). Reconceptualizing Political Accountability. *International Political Science Review / Revue internationale de science politique*, *19*(4), 387-406.

Moyo, D. (2009). *Dead aid: Why aid is not working and how there is a better way for Africa*. Macmillan.

Murakami, H (1989). *A Wild Sheep Chase.* New York: Random House

Murakami, H. (2019). *Killing Commendatore: A novel*. New York: Vintage International.

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies (Qualitative Research Methods Vol. 11).* Thousand Oaks: SAGE. Available from: https://dx.doi.org/10.4135/9781412985000

OECD. (2005). *Paris Declaration on Aid Effectiveness*. Paris: OECD https://doi.org/doi:https://doi.org/10.1787/9789264098084-en

OECD. (2008). *Accra Agenda for Action*. Paris: OECD https://doi.org/doi:https://doi.org/10.1787/9789264098107-en

OECD. (2010). *Evaluating Development Co-operation. Summary of key norms and standards.* Paris: OECD Publishing.

Oosterom, M., & Scott-Villiers, P. (2016). Power, poverty and inequality. *IDS Bulletin*, *47*(5). https://doi.org/10.19088/1968-2016.161

Ottoson, J. M. (2009). 'Knowledge-for-action theories in evaluation: Knowledge utilization, diffusion, implementation, transfer, and translation'. *New Directions for Evaluation, 2009(*124), 7-20.

Owen, J. M. (2007). *Program evaluation: Forms and Approaches* (3rd ed.). New York: Guilford Press.

Packard Foundation. (2010). *Linking Knowledge with Action*. Retrieved from Los Altos: https://www.packard.org/insights/resource/linking-knowledge-with-action-an-approach-to-philanthropic-funding-of-science-for-conservation/, accessed 14 July 2020

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). London: SAGE.

Patton, M. Q. (2010). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford Press.

Patton, M. Q. (2014). *Qualitative Research & Evaluation Methods: Integrating theory and practice*. Thousand Oaks: SAGE.

Patton, M. Q. (2019). *Blue Marble Evaluation: Premises and Principles.* New York: Guilford Press.

Pawson, R. (2000). 'Middle-range realism'. *European Journal of Sociology / Archives Européennes de Sociologie / Europäisches Archiv für Soziologie, 41*(2), 283-325.

Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. Thousand Oaks: SAGE.

Peruzzotti, E. (2014). 'Accountability Deficits of Delegative Democracy'. In D. Brinks, M, Leiras, & S. Mainwaring (Eds.) *Reflections on Uneven Democracies: The Legacy of Guillermo O'Donnell*. Baltimore: Johns Hopkins University Press.

Picciotto, R. (1995). 'Introduction: Evaluation and development'. *New Directions for Evaluation, 1995*(67), 13-23.

Picciotto, R. (2002). *The Logic of Renewal: Evaluation at the World Bank (1992-2002)*. Washington, DC: World Bank. Available from: https://openknowledge.worldbank.org/handle/10986/20228?locale-attribute=en

Picciotto, R. (2003). 'International Trends and Development Evaluation: The Need for Ideas'. *American Journal of Evaluation, 24*(2), 227-234. Available from: https://doi:10.1177/109821400302400206

Picciotto, R. (2018). 'Accountability and learning in development evaluation: A commentary on Lauren Kogen's thesis'. *Evaluation, 24*(3), 363-371. Available from: https://doi:10.1177/1356389018781361

Picciotto, R., & Wiesner, E. (1998). Evaluation and Development: The Institutional Dimension (Vol. 1). New Brunswick, NJ: Transaction Publishers.

Pilkington, H. (2018). 'Employing meta-ethnography in the analysis of qualitative data sets on youth activism: a new tool for transnational research projects?' *Qualitative Research, 18*(1), 108-130. Available from: https://doi:10.1177/1468794117707805

Plaatjies, D., & Porter, S. (2011). 'Delivering on the Promise of Performance Monitoring and Evaluation'. In D. Plaatjies (Ed.), *The Future Inheritance: Building State Capacity in Democratic South Africa*. Johannesburg: Jacana.

Pollitt, C., Bouckaert, G., & van Dooren, W. (2009). *Measuring Government Activity*. Paris: OECD.

Porter, S. (2011). ''Helping' as an Evaluation Capacity Development Strategy in South Africa'. In R. C. Rist, M-H. Boily, & F. Martin (Eds.), *Influencing change: Building Evaluation Capacity to Strengthen Governance.* Washington, DC: World Bank.

Porter, S. (2012). 'Moving Beyond Teaching People to Fish: VOPE's Role in Sustainable Learning Strategies by Working as Innovation Intermediaries'. In M. Segone & J. Rugh (Eds.), *Evaluation and Civil Society: Stakeholders' perspectives on National Evaluation Capacity Development*. New York: UNICEF.

Porter, S. (2013). 'A Change of road for a rights-based approach? A case study of mobile technology in KwaZulu Natal, South Africa'. In P. Gready & W. Vandenhole. (Eds.), *Human rights and development in the new Millennium: Towards a theory of change*. London: Routledge.

Porter, S. (2016). 'A framework for improving the responsiveness of policy through development evaluation'. *Evaluation, 22*(2), 245-258. Available from: https://doi:10.1177/1356389016638623

Porter, S. (2017). 'A Framework for Identifying Entry Points for International Development Evaluation to Enable Responsive Government Policy'. In D. Podems (Ed.), *Democratic Evaluation and Democracy: Exploring the Reality*. Charlotte, NC: Information Age Publishing.

Porter, S., & de Wet, J. (2009). Who will guard the guardians? Amartya Sen's contribution to development evaluation. Development in Practice, 19(3), 288-299.

Porter, S., Djidjoho, A. N., Houinsa, D. G., Amoatey, C., Machuka, S., Okumu, B. W., . . . Asingwire, N. (2012). *African Monitoring and Evaluation Systems: Exploratory Case Studies*. Johannesburg: Graduate School of Public and Development Management, University of the Witwatersrand. Available from: http://wiredspace.wits.ac.za/handle/10539/28799, accessed on: 9 February 2020

Porter, S., & Gasana, C. (2013). *Study on the Demand for and Supply of Evaluation in Rwanda*. Johannesburg: Graduate School of Public and Development Management, University of the Witwatersrand. Available from: http://wiredspace.wits.ac.za/handle/10539/28235, accessed on: 9 February 2020

Porter, S., & Goldman, I. (2013). 'A Growing Demand for Monitoring and Evaluation in Africa'. *African Evaluation Journal, 1*(1), 9 pages. Available from: https://doi.org/10.4102/aej.v1i1.25

Porter, S., & Hawkins, P. (2019). 'Achieving Sustainability through Sustainable Organizational Evaluation Systems'. *New Directions for Evaluation, 2019*(162), 87-101. Available from: https://doi.org/10.1002/ev.20360

Porter, S., & Mulenga, O. (2013). *Study on the Demand for and Supply of Evaluation in Zambia*. Johannesburg: Graduate School of Public and Development Management, University of the Witwatersrand. Available from: http://wiredspace.wits.ac.za/handle/10539/28236, accessed on: 9 February 2020

Preskill, H., & Boyle, S. (2008). 'A Multidisciplinary Model of Evaluation Capacity Building'. *American Journal of Evaluation, 29*(4), 443-459. Available from: https://doi.org/10.1177/1098214008324182

Przeworski, A., Stokes, S. C. S., Stokes, S. C., & Manin, B. (1999). *Democracy, accountability, and representation* (Vol. 2). Cambridge University Press.

Raftree (2013) 'Poverty Porn: a reflection of deeper issues' *Wait…What? Discussions on digital ethics. Privacy and Power*. Available from: https://lindaraftree.com/category/poverty-porn/, accessed on: 6 Febraury 2021

Reeler, D., Van Blerk, R., Taylor, J., Paulsen, D., & Soal, S. (2009). *Barefoot Guide to Working with Organisations and Social Change*. Cape Town: CDRA.

Robeyns, I. (2005). 'The capability approach: a theoretical survey'. *Journal of human development, 6*(1), 93-117.

Roche, C. (2015). 'The politics of juggling multiple accountability disorder' in Eyben, R., Guijt, I., Roche, C., & Shutt, C. (Eds.). *The politics of evidence and results in international development: Playing the game to change the rules?*. Rugby: Practical Action Publishing.

Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2018). *Evaluation: A systematic approach*. Thousand Oaks: SAGE.

Runciman, D. (2018). *Political Hypocrisy: The Mask of Power, from Hobbes to Orwell and Beyond*, Revised Edition. Princeton, NJ: Princeton University Press.

Rutman, L. (1990). 'Serving Management and Accountability: A Conflict for Program Evaluation?' *Evaluation Practice, 11*(1), 39-43. Available from: https://doi.org/10.1177/109821409001100106

Ryan, K. E. (2002). 'Shaping educational accountability systems'. *American Journal of Evaluation, 23(*4), 453-468.

Ryan, K. E. (2004). 'Serving public interests in educational accountability: Alternative approaches to democratic evaluation'. *American Journal of Evaluation, 25*(4), 443-460.

Ryan, K. E. (2005). 'Making educational accountability more democratic'. *American Journal of Evaluation, 26*(4), 532-543.

Ryan, K. E., Geissler, B., & Knell, S. (1996). 'Progress and accountability in family literacy: Lessons from a collaborative approach'. *Evaluation and Program Planning, 19*(3), 263-272.

Schein, E. (2009). *Helping: How to Offer, Give, and Receive Help*. San Francisco: Berrett-Koehler Pub.

Schillemans, T. (2011). 'Does Horizontal Accountability Work?: Evaluating Potential Remedies for the Accountability Deficit of Agencies'. *Administration & Society, 43*(4), 387-416. Available from: https://doi.org/10.1177/0095399711412931

Schmitter, P. C. (2004). 'The ambiguous virtues of accountability'. *Journal of Democracy, 15(*4), 47-60.

Schoenefeld, J. J., & Jordan, A. J. (2019). 'Environmental policy evaluation in the EU: between learning, accountability, and political opportunities?' *Environmental Politics, 28*(2), 365-384. Available from: https://doi.org/10.1080/09644016.2019.1549782

Scriven, M. (1981). *The Logic of Evaluation*. Claremont: Edgepress.

Scriven, M. (1991). *Evaluation Thesaurus* (4th ed.). Thousand Oaks: SAGE.

Scriven, M. (1995). The logic of evaluation and evaluation practice'. *New Directions for Program Evaluation, 1995*(68), 49-70.

Sen, A. K. (1992). *Inequality Reexamined.* New York: Russell Sage Foundation Books at Harvard University Press.

Sen, A. K. (1999). *Development as Freedom*. Oxford: Oxford University Press.

Sen, A. K. (2005). 'Human Rights and Capabilities'. Journal of Human Development, 6(2), 151-166.

Sen, A. K. (2007). *Identity and Violence: The Illusion of Destiny*. London: Penguin.

Sen, A. K. (2009). *The Idea of Justice*. Cambridge, MA: Harvard University Press.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). 'Foundations of Program Evaluation: Theories of Practice'. Thousand Oaks: SAGE.

Shah, A. (1998). 'Fostering fiscally responsive and accountable governance: Lessons from decentralization'. In R. Picciotto & E. Wiesner (Eds.), *Evaluation and Development: The Institutional Dimension* (pp. 81-107). New Brunswick, NJ: Transaction Publishers.

Stame, N. (2003). 'Evaluation and the policy context: the European experience'. *Evaluation Journal of Australasia, 3*(2), 36-43.

Stiglitz, J. E. (2002). *Globalization and its Discontents*. New York Norton.

Stiglitz, J. E. (2002). 'Information and the Change in the Paradigm in Economics'. *The American Economic Review, 92*(3), 460-501.

Sørensen, E., & Torfing, J. (2016). *Theories of Democratic Network Governance.* London: Palgrave Macmillan.

Taylor-Powell, E., & Boyd, H. H. (2008). 'Evaluation capacity building in complex organizations'. *New Directions for Evaluation, 2008*(120), 55-69.

Tortajada, C. (2010). Water Governance: Some Critical Issues. *International Journal of Water Resources Development*, *26*(2), 297-307. https://doi.org/10.1080/07900621003683298

Tuhiwai Smith, L. (2018). Indigenous insight on valuing complexity, sustaining relationships, being accountable. *Tackling wicked problems in complex ecologies: the role of evaluation. Stanford University Press, Palo Alto, CA*, 45-66.

UN General Assembly. (2015). *Transforming Our World: The 2030 Agenda for Sustainable Development*. General Assembly 70 session. 21 October 2015, A/RES/70/1. Available from: https://www.refworld.org/docid/57b6e3e44.html, accessed on: 14 July 2020

Van Der Meer, F.-B., & Edelenbos, J. (2006). 'Evaluation in multi-actor policy processes: accountability, learning and co-operation'. *Evaluation, 12*(2), 201-218.

VandenBos, G. R. (Ed.) (2013). *APA Dictionary of Clinical Psychology.* American Psychological Association. Available from: https://doi.org/10.1037/13945-000

Van Dooren, W., Pollitt, C., & Bouckaert, G. (2009). *Measuring Government Activity*. Paris: OECD Publishing.

VSO-RAISA, (2007). Challenges of Care. *VSO-RAISA Regional Conference Report Back Pretoria, South Africa, 1 - 2 November 2007*. Pretoria: VSO-RAISA. Available from: http://www.bibalex.org/Search4Dev/files/288485/119529.pdf, accessed on: 23 May 2020

Waterman, R. W. and Meier K. J. (1998) Principal-Agent Models: An Expansion?, *Journal of Public Administration Research and Theory*, Volume 8, Issue 2, April 1998, Pages 173–202, https://doi.org/10.1093/oxfordjournals.jpart.a024377

Weaver, C. (2008). *Hypocrisy Trap: The World Bank and the Poverty of Reform*. Princeton, NJ: Princeton University Press.

Weiss, C. H. (1993). Where Politics and Evaluation Research Meet. *Evaluation Practice*, *14*(1), 93-106. https://doi.org/10.1177/109821409301400119

Weiss, C. H. (1998). 'Have We Learned Anything New About the Use of Evaluation?' *American Journal of Evaluation, 19*(1), 21-33. Available from: https://doi.org/10.1177/109821409801900103

Wiesner, E. (2011). 'The Evaluation of Macroeconomic Institutional Arrangements in Latin America'. In R. C. Rist, M-H. Boily, & F. Martin (Eds.), *Influencing change: evaluation and capacity building driving good practice in development and evaluation*. Washington: World Bank.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users (*3rd ed.). Thousand Oaks: SAGE.

Zapico-Goñi, E. (2007). Matching public management, accountability and evaluation in uncertain contexts: a practical suggestion. *Evaluation, 13*(4), 421-438.

Appendix A: Porter, S., & de Wet, J. (2009). Who will guard the guardians? Amartya Sen's contribution to development evaluation. Development in Practice, 19(3), 288-299.

Development in Practice, Volume 19, Number 3, May 2009

Routledge
Taylor & Francis Group

Who will guard the guardians? Amartya Sen's contribution to development evaluation

Stephen Porter and Jacques de Wet

An Action Learning process integrated with Sen's Capability Approach can support development agencies to formulate interventions that enhance freedom. The authors show that putting this approach into practice has important implications for the manner in which 'development' is undertaken as an ideological project. It may help to examine and challenge those who hold power in development: the guardians. This finding is the result of an emergent Action Learning process that was initiated by applying Sen's principles to focus-group interviews with women who care for people affected by HIV and AIDS. One of the findings of these focus groups was that the participants valued the process because it opened a space for them to influence the work of the implementing NGO. Essentially, they could hold the implementing agency to account. Reflection on this outcome by the agency led to important shifts in processes that are more supportive of freedom.

KEY WORDS: Civil Society; Methods; Sub-Saharan Africa

Introduction

This article is written from the perspective of the struggle of practitioners in planning, monitoring, and evaluation (PME) to translate relevant theory into everyday practice. We are looking not only for tools to aid participatory evaluation, but also for concepts that enable us to understand how development work is constrained by narrow limits prescribed by donor and development-implementing agencies, which can promote the interests of right-wing-dominated global systems of power. This article does not simply represent a case study of how one NGO works with local partners to enhance impact, but reflects on how an implementing agency, working with the concept of development as freedom, supports community development work in an unpredictable and changing environment.

We argue that development agencies can better support freedom through their interventions by opening spaces for Action Learning processes informed by Amartya Sen's Capability Approach. This is the key finding that has emerged from a continuing Action Learning

ISSN 0961-4524 Print/ISSN 1364-9213 Online 030288-12 © 2009 Oxfam GB
DOI: 10.1080/09614520902807987 Routledge Publishing

process facilitated by PME practitioners in the Voluntary Services Overseas Regional AIDS Initiative of Southern Africa (VSO–RAISA).[1] This argument may not appear particularly radical, yet the consequences of putting into operation an Action Learning cycle that is informed by Sen's Capability Approach has implications not only in a direct way for programme learning, but for an agency's ability to challenge the ideological basis of the 'development' project. It may help in examining and challenging the roles of guardians, such as implementing agencies and donors, who use their power in a manner that affects what development can achieve.

The work of Sabina Alkire, which is grounded in Sen's Capability Approach, alerts us to the importance of viewing community-level development from the perspective of development as freedom, which is the ethical priority of the Capability Approach. 'Freedom' here refers to the capability of a person to choose to enjoy or to disregard life choices that expand human well-being, such as being well-nourished or avoiding escapable morbidity. Further, in viewing development as freedom you do not aim only for an end state: you ensure that how you get there is consistent. In short, you ensure that people are not forced to be free. Alkire (2002) not only provides an accessible overview of the Capability Approach, but also a focus-group methodology which facilitates the measurement of desirable and undesirable changes in people's lives. This methodology was used in Zambia and South Africa with volunteer home-based carers of people affected by HIV and AIDS. The focus-group discussions provided the care givers with an opportunity to influence the approach of VSO–RAISA: their advice – 'first you take care of the care giver, then you take care of the client' – has directly fed into practice. The care givers also valued the focus-group process itself, because it enabled participants to develop relationships with one another, and to realise that they were experiencing similar struggles and achievements.

These insights enabled the PME practitioners to reflect critically upon how the nascent Action Learning processes of VSO–RAISA could be strengthened by Sen's Capability Approach. Sen (1999: 3, 36) sees development as a process 'of expanding the real freedoms that people enjoy'. In his discussion of the attainment of freedom, Sen (1999: 4) emphasises two issues: that 'the achievement of development is thoroughly dependent on the free agency of people'; and that 'the productive role of public discussion, social agitation, and open debates' (Sen 2005: 160) is a central part of the mechanism of evaluation.

Action Learning itself is the constant conscious process of moving from doing, to reflection, to thinking, to improving, and then back to doing. Action Learning has been summarised neatly by Ortrum Zuber-Skerrit (Ortrum Zuber-Skerrit, quoted in CDRA 2005: 1):

> *Action Learning is learning from concrete experience, through group discussion, trial and error, discovery and learning from and with each other. It is a process by which groups of managers/leaders or 'learners' generally work on real issues or problems, carrying real responsibility in real conditions.*

The implementation of an Action Learning process informed by the view of development as freedom provides us with a means to appreciate more clearly the relevance of discussions with the care givers within and beyond VSO–RAISA programmes, because it directs us to the broader concern of how 'development' is undertaken as an ideological project. Sen's Capability Approach helps us to examine the role of the guardians of the public good by directing our attention towards the need for open debate and informed scrutiny. The maxim '*Quis custodiet ipsos custodes?*' (Who will guard the guardians?) (we discuss the identity of the guardians below) is a matter of central importance in political science, and an appropriate principle to guide us in our critical reflections on development practice.

This article is structured like an hourglass. First, we introduce the meta-issues, which draw attention to the reality that evaluation practice, together with those values that direct it, often limits the scope of development. Second, we introduce the Capability Approach and its relation to these issues. Third, we discuss how the results of the focus groups directed attention to the role of Sen's Capability Approach in challenging guardians within the development arena. Fourth, we return to the meta-issues. Finally, by way of concluding remarks, we introduce some questions which explore the extent to which the guardians of development can be guarded.

The meta-issues

In any development project the donors, the implementing agencies, and the target group all have their own agendas, as well as the ability to manipulate development processes to their own ends (although they are not necessarily equally powerful): they are all guardians in one sense or another. Donors hold the purse strings and set the terms of projects and, in many instances, define the criteria for success. Implementing agencies produce funding proposals whereby those projects that they perceive to be valuable can be implemented with the assistance of project partners. Implementing agencies then collect data in order to secure funding for future projects. Project partners, local community organisations, and the target groups – those whom the project is supposed to benefit – are able to manipulate the collection of data (for example by concealing information from evaluators) and so subvert a project to suit their own ends.

With few donors, many implementing agencies, and target groups of millions, power is concentrated at the top of the aid pyramid. The values that shape development projects are to a large extent defined by donors and implementing agencies.[2] '[V]alues lie at the heart of evaluation' (House 2005:1072), either explicitly or implicitly. These values directly determine what is to be monitored and evaluated in current projects, and they also influence what projects will be undertaken in the future. As an evaluation method, Action Learning is, therefore, among other things, an exercise of power. In defining, executing, and analysing a project, the sometimes-myopic values of stakeholders (whether donors, implementing agencies, or target groups) determine what is taken as relevant. Although PME practitioners are to some extent able to influence the process – for example, through the information produced – there are limits to their power. If they adhere rigidly to guidelines that implicitly or explicitly favour the values of donors and implementing agencies only, they prescribe and limit what can be taken as relevant. In practice, the values of donors largely define what is to be monitored and evaluated. For example, they can circumscribe to a greater or lesser extent how those at the lower levels of the aid structures are to be engaged with, and to what degree they are able to challenge or adjust the development process to coincide with their own values.

Two examples illustrate this point. Our first example is about the influence exerted on PME practices through certain recent changes in the political landscape that can be seen as part of the 'War on Terror'. Denzin and Lincoln (2005: 9) refer to 'Bush Science' in which 'experimental, evidence-based methodologies represent a racialized, masculinist backlash to the proliferation of qualitative enquiry over the past two decades'. In a related example, House (2005: 1078–9) demonstrates the manner in which classic control-group methodologies have been prescribed by law in the USA, to the exclusion of other lines of enquiry or knowledge. The result is that indicators are defined solely by Washington, according to a predetermined methodological fundamentalism that is aligned with certain world views and values; and not in conjunction with stakeholders at a community level (Denzin and Lincoln 2005: 8–10; House 2005: 1078–80). Whether or not one agrees with the characterisation of 'Bush Science', there is weight in the argument that the PME parameters of US initiatives present narrowly defined indicators, which are shaped by a conservative liberal capitalist ideology.

Evidence of this approach to development can be seen in the President's Emergency Plan For AIDS Relief (PEPFAR), promoted by the former US President George W. Bush. For example, the PEPFAR Act (US Congress 2003: 117, Stat 713) clearly gives a special role to faith-based organisations that 'possess infrastructure, experience, and knowledge that will be needed to carry out these [prevention, care and support] programs in the future and should be an integral part of United States efforts'. Further, although social and behavioural causes of HIV are mentioned, the Act states 'HIV/AIDS is first and foremost a health problem'. Bush's religious beliefs and narrow health-oriented perspective on the pandemic are central to the PEPFAR vision, while the socio-economic roots of the pandemic are marginalised. The consequences of this are apparent in the strategic plan and indicators. For example, the strategic plan limits condom distribution in a wide variety of settings (PEPFAR 2006), while a relatively narrow set of quantifiable indicators has been defined and used to judge the success of the strategic plan. Regarding the indicators, PEPFAR (2003:73) states:

> *Core indicators will include outcome indicators addressing the Emergency Plan's '2-7-10' goals of 2 million people treated, 7 million infections prevented, and 10 million persons provided with care. Program progress indicators include numbers of facilities supported, numbers of practicing professionals and community workers trained, and numbers of clients reached through prevention, treatment, and care interventions.*

Aside from the problematic nature of trying to measure 'infections prevented', these indicators allow the target group only very limited space to participate in the process of evaluation. In the PEPFAR approach, all that matters is the size of the target group, which is simply to be counted, while values that they may hold dear are systematically excluded in any evaluation of how resources are distributed. Corroborating this statement are concerns aired in a number of forums attended by one of the present authors. For example, at a recent meeting of Southern African regional NGOs it was observed that psychosocial support for AIDS orphans is being neglected by partner organisations because – since they accept PEPFAR funding – they are compelled to concentrate on types of service provision that are easily quantified. Distributing books is preferred and treated as synonymous with psychosocial support, rather than the long-term counselling of traumatised children, which is the method preferred by a number of regional NGOs who work to support vulnerable children. It is not that the provision of books is unimportant; rather that a suite of interventions which recognise the children's psychosocial and educational needs and rights is central to mitigating the impact of AIDS.

Our second example draws attention to the fact that even an implementing agency which emphasises rights, participation, and empowerment can favour those objectives related to their own ideological stance. This can lead to project evaluations that obscure changes felt by the community. For example, a study of ActionAid's *Reflect* methodology revealed that, despite changing their language in development interventions and evaluation to incorporate a Freirean approach, the evaluations recreated narratives and situations that would be familiar to colonial administrators and missionaries (Fiedrich and Jellema 2003: vii–viii). It was found that ActionAid representatives would often simplify stories of change in order to meet predetermined criteria. The women involved in the programme, meanwhile, would tell evaluators what they thought they wanted to hear. On occasion the women would use the project to serve their own ends, which often were contrary to those of ActionAid, although this would not arise in evaluation reports (Fiedrich and Jellema 2003: 28–33). It could be argued that the target group was legitimately influencing the distribution of resources via the evaluations, yet the evaluation lacked open, two-way lines of accountability despite its participatory nature. Adherence to restricted language and a narrow informational base limited what the ActionAid evaluations would find. The consequences were twofold. First, except for those

who manipulated the process, most members of the target group were unable to incorporate their own values into project design. Second, because of this manipulation, violations of human rights could occur, unwittingly abetted by development interventions.

Two lessons emerge from reviewing the above situation. First, in working with implementing agencies, it is of central importance for members of the target group to discuss their values and to be able to influence the organisational value system, rather than having to resort to the above-mentioned manipulation. Second, in its PME system, the implementing agency should be flexible enough to enable the voices and values of those within targeted communities to be heard, both so that its operations do not encroach on the freedoms of disadvantaged minorities, and also to better inform future practice. As shown by the examples of PEPFAR and ActionAid, PME systems can cultivate their own 'truth', which may be quite different from the reality experienced by the target group.[3] This is fraudulent misrepresentation of the impact of a project upon people: projects designed to alleviate poverty are judged successful even though they have not enhanced people's lives. On the other hand, 'successful projects' are judged as failures (Sen 1992: 30; Fiedrich and Jellema 2003: 76–8).

Development interventions have inflicted suffering and violence upon the poorest members of society because they are distorted by a world system which favours the rich and powerful, and they are conceptually flawed by this bias, which ignores the views and values of the target group (Gasper 2004: 9–11; Sen 1992: 28–30, 1999: 41–3). An approach that encourages critical reflection within an implementing agency allows greater freedom in what PME can find, and can promote the questioning of the values and priorities of those in positions of power in development.

The Capability Approach and its potential to help to guard the guardians[4]

In our study we argue that an Action Learning process informed by the Capability Approach supports freedom by helping to keep open a space for the examination of the issue of guarding the guardians. The Capability Approach has a number of tools that can be incorporated into the Action Learning cycle, and in addition allows critical reflection on how a programme supports the achievement of freedom. We argue that by giving a central position to public discussion, social agitation, and open debate, and understanding development as freedom, a wider array of stakeholders can exercise power in a more open and explicit manner. As Sen (1999: 9) has stated: 'The exercise of freedom is mediated by values, but the values in turn are influenced by public discussions and social interactions, which are themselves influenced by participatory freedoms'. In our study, Action Learning is mediated by values, which in turn ought to be influenced by public discussions that are conducted in a manner which ensures that all parties participate freely – that is, without their voices being subverted or suppressed. Two of the more theoretical aspects of the Capability Approach are particularly supportive of this: its interaction with other theories and its ethical priorities.

The manner in which Sen's Capability Approach can be accommodated in a number of alternative approaches strengthens mutual accountability in PME practice. The merits of various competing 'development' approaches, such as utilitarianism (Sen 1992: 55; Sen 1999: 62–3), the Basic Needs Approach (Alkire 2002: 15, 170), and the rights-based approach (Sen 2005) are acknowledged in the Capability Approach and used to enrich it. Herein lies the value of the Capability Approach in aiding evaluation: its ability to offer an enhanced approach to Action Learning within the existing limits of knowledge. In short, this enables evaluation to take place within a broad spectrum of existing ideas and informational criteria.

The ethical priority of the Capability Approach is the promotion of freedom. Development is considered to be the removal of ills, such as poverty and tyranny, and the enhancement of valued ways of 'doing and being', for example, literacy and political participation. Sen (1999: 4) has stated:

Freedom is the central process of development for two distinct reasons.

1 *The evaluative reason: assessment of progress has to be done primarily in terms of whether the freedoms that people have are enhanced;*
2 *The effectiveness reason: achievement of development is thoroughly dependent on the free agency of people.*

In short, freedom is both the '*primary end* and . . . the *principal means* of development' (*ibid.*: 36, original emphasis).

What this ethical stance means in practice is that focusing on the *consequences* of development is in itself insufficient: *how* the developmental process is achieved is also of great importance. In Soviet Russia, for example, it could be imagined that people were able to live in reasonable housing and their children have access to schooling, but yet they would not have the freedom to move jobs within the labour market or worship as they might choose. The centrality of freedom as both *means* and *ends* has been emphasised throughout Sen's major writings (see for example Sen 1992; 1999). He contrasts this sharply with more economic approaches that place commodities at the centre of development.

Sen's ethical stance demands that a broad array of information should be taken into account in order to understand the manner and consequences in which development takes place. For Sen (1999: 57) 'the real "bite" of a theory of justice can, to a great extent, be understood from its informational base: what information is – or is not – taken to be directly relevant'. In the same sense, the 'bite' of an evaluation framework can be understood by what the framework does – or does not – take to be relevant. For example, when eating a lemon meringue pie, your enjoyment of its flavour would be conditioned by the flavour of each spoonful – whether lemon, meringue, or both. If you change the spoon for a larger one, your appreciation of the pie and its components changes. Changing a framework and the values that underlie an Action Learning cycle can help one to take a different perspective and clarify the components under evaluation. The Capability Approach does not place limits on the information taken to be relevant, and hence it contrasts with the earlier examples of PEPFAR and ActionAid. In these examples, the informational base was deliberately limited so that the evaluation was narrowly focused, consequently making it difficult for those with power to be held accountable. It is not that we always need the biggest spoon to gather all information, more that each case analysed should be sensitive and open to inputs from all parties.

It is also important to note that the ethical priority for a capability-based Action Learning cycle is to evaluate a development intervention not solely in terms of how it enhances freedom, but to ask also if it does so in an ethically consistent manner. This means that in the Capability Approach the Action Learning cycle must work in ways that do not limit our understanding of freedom and do not encourage the promotion of servitude. Gasper (2002: 176) has summarised this position well: 'The normative priority to capability can be read as a policy rule to promote people's capability and then "let them make their own mistakes", rather than of achieved well-being or quality of life, though it might contribute to them'. And Sen (2005: 157) has stated that the 'richness of the capability perspective broadly interpreted . . . includes its insistence on the need for open scrutiny for making social judgements'. Essentially, in the operation of an Action Learning cycle informed by the Capability Approach we should aim for an open discussion of programme development within the implementing agency, and with partner organisations, and with target groups.

From the above it can be seen that the Capability Approach can contribute to Action Learning in organisations in two ways: first, the Capability Approach adapts and works with other approaches; as Gasper (2004: 158) has argued: 'the supposedly rival concepts of basic needs, capabilities, and human rights . . . are not really rival concepts at all, but are essentially complementary'. This enables an implementing agency to draw on a broad range of approaches, with an incomplete informational basis. Second, the ethical priority of the Capability Approach demands that the *means* as well as the *ends* of an intervention are consonant with freedom. In practice this requires an implementing agency to include processes that enhance freedom, for example public discussion of values, and making provision for participants to hold those in power accountable.

An Action Learning cycle informed by the Capability Approach: the story so far

This section deals with the actual experience of implementing an Action Learning cycle, informed by the Capability Approach, which represents reflection on an emergent set of events. Looking back on the process thus far has been like trying to solve a problem without knowing what the specific components are: if the Capability Approach looks like X in theory, how can this apply to broader programming, Y, in a given organisational context, Z, and in relation to a world system that limits the basis of evaluation, A?

In the first instance, the investigation of Sen's Capability Approach was undertaken to see how the approach could be put into operation to enhance the PME of a development programme. On applying Sen's ethic of freedom to VSO–RAISA's PME framework it was apparent that not enough attention had been given to allowing people to be the 'authors of their own destiny'. VSO–RAISA's PME system does report on valuable changes in people's lives, yet it did not specifically enable them to participate in evaluating interventions and challenging VSO–RAISA on the direction of future interventions.

In an attempt to address this shortcoming, a focus-group methodology, grounded in Alkire's (2002) use of Sen's Capability Approach, was adapted and implemented with two groups of home-based care givers in Zambia and South Africa. The care givers had been touched by VSO–RAISA directly through a variety of interventions that affected their lives: for example, through training in treatment literacy and through the provision of small grants to help them undertake income-generating activities. Alkire's focus-group methodology seeks to enable people to enter into open public discussion: participants are given the space to recognise, define, and choose instances which help or hinder them. The open-ended nature of the methodology allows for the exploration of the intended and unintended effects of an intervention upon a target group. Participants can express the manner in which an intervention had altered their lives favourably or unfavourably and can influence future development interventions through reference to their own objectives.[5]

The focus group does not operate in a vacuum. A framework exists to help facilitators to probe issues that arise during the process. Alkire (2002) derived eight dimensions of impact from the religious philosopher John Finnis. These dimensions, we argue, are central to the methodology. They enable the facilitator of the focus group to support appreciative inquiry into changes, while also sensitively drawing attention to struggles and challenges in certain dimensions. For example, when discussing relationships, the challenges regarding gender relations in implementation can be probed. Alkire (2002: 52) asserts that the dimensions of impact (see Table 1) are

Table 1: Alkire's Eight Dimensions of Impact

Dimension of impact	Description
Life–health–security	Changes related to physical survival
Knowledge	Knowing: technical, practical, about others, and about themselves
Excellence in work and play	Impact on skills used at work, and at home during relaxation
Relationships	Within community, family, with outsiders, within groups, between men and women
Inner voice	At peace with themselves, with their conscience, sense of harmony
Empowerment	Ability to make meaningful choices and decisions and to influence others
Beauty/environment	Impact on environment, sense of harmony with nature: has intervention created or destroyed things of beauty?
Religion	Impact on deeper values, sources of meaning

Source: Alkire 2002: *267–71*

> *like the 'primary colours' of values. An infinite range of shades can be made from our three primary colours, and not every painting (or life or community or income generation project) uses all or even most shades, but if, for example, all red hues were entirely missing, then my understanding of colour would be consistently skewed.*

Alkire stresses that this 'list' is by no means complete. It offers an implementing agency a useful entry point to comprehend values within a broad informational basis. The dimensions are defined in Table 1.

Participants in the two focus groups were able to challenge the values of the implementing agency, VSO–RAISA, because they were able to discuss the interventions in relation to the lives that they valued, exercising a form of public discussion. A statement from the Zambian participants movingly summarises the majority of recommendations of both groups of care givers during the focus groups: 'First you take care of the care giver, then you take care of the client'. This process, it can be argued, is supportive of creating valued freedom, as it helped VSO–RAISA to redefine interventions related to care givers, while creating space for two-way accountability: for example, care for the care givers by VSO–RAISA, supporting the development of volunteer management systems.

Useful as the focus groups were in informing focus on care for the care givers, by itself the process can only give a very limited snapshot of change. These focus groups were *ex-post* analysis without corresponding *ex-ante* studies. This limits their range in terms of more broadly explaining change. Although further investigation was undertaken to see if the focus-group methodology could be extended, it was found that because of the constraints faced by field staff only a limited number of related implementations of the focus-group methodology was possible.

Yet, the effects of the focus group have been felt beyond its implementation and analysis. Most importantly the ethics of the Capability Approach came to inform VSO–RAISA's Action Learning cycle and how it relates to project partners and target groups. Supporting the Action Learning cycle in guarding the guardians was realised chiefly by placing emphasis on open discussion and taking care that limits on what information was considered relevant were not unthinkingly imposed. Within VSO–RAISA, informing the Action Learning cycle with Sen's Capability Approach found resonance with the values held by many of the programme staff. Although the Capability Approach in its entire form was not introduced,

the methods, language, and tools introduced in dialogue supported programme staff in implementing their work in a way that reflected their existing values.

By modifying some of processes of the Action Learning cycle as a result of the focus groups, and treating participation and open discussion as ends in themselves, different avenues have been opened. For example, there have been discussions of what PEPFAR money can mean, and how VSO–RAISA can support partners who receive this money. This has led to support for partners in maintaining their own vision and values when receiving PEPFAR money, and the inevitable array of quantitative indicators that come with it. This is evidence that the Capability Approach has influenced recognition of issues in the meta-environment, and supported work with project partners and target groups that in a small way challenges some of those issues.

Community organisations have consistently reported the value of these processes. Five community organisations (themselves made up of target groups) from across Southern Africa have commented directly on the intrinsic value of the process in interviews with one of the authors, who is part of VSO–RAISA. Similar processes have been implemented with approximately 26 organisations. Anecdotal evidence suggests that all current staff have affirmed the value of the process. Finally, a donor evaluation, which focused upon participation within the programme, cited the PME system as an example of best practice. These affirmations mean that the ethics of the Capability Approach have found actuality in both the *means* and the *ends* in the Action Learning cycle of VSO–RAISA; this illustrates a broader change in thinking and approach within VSO–RAISA. Alkire's framework has also informed qualitative programme indicators and been used in a training workshop held with partners.

As noted at the start of this section, this way of working is continuing and is still encountering challenges internally and externally. The challenges of the meta-environment, where what can be imagined is restricted, can still directly affect the Action Learning cycle. For example, externally the development sector continues to work in a fragmented form with different goals and separate projects. The Millennium Development Goals (MDGs) contribute to this fragmentation. Communities are whole entities that cannot be segmented in this manner. When applying for new funding, one often has to speak to these goals and so unwittingly reduce the range of work possible. Internally, other management processes, such as budgeting and decisions relating to programme priorities, can have a positive or an adverse impact upon the current implementation of the Action Learning cycle. VSO–RAISA's work spans six countries, with the Action Learning cycle moving at different speeds, with varying levels of success. All of these challenges take place in a context where introducing complex concepts in a succinct and meaningful manner and the high workload of programme staff compound the difficulties.

Over the course of the year, VSO–RAISA has been exploring approaches based on human rights. (In brief, a human-rights approach links national and international legal instruments to current development programming, to better enable people to claim their rights. This current exploration has led to an appreciation of the number and type of instruments available and the duties of governments to move towards realisation of these rights.) The language and concepts of the human-rights approach complement the Capability Approach. The aim of this exploration is to develop a framework that is recognised internationally and accepted within the development community, and which also speaks to the internal and external issues noted above. Especially important in this respect is the recognition of the inalienability, interdependence, and interconnectedness of rights. This stands in contrast to the more fragmented approaches currently prevalent in development work.

The Capability Approach, because of its ethical focus on the *means* as well as the *ends* of development, encouraged VSO–RAISA to reflect upon its values through opening its Action Learning cycle to project partners *and* target groups. This has supported development as freedom by enabling mutual accountability; and it has brought changes in programme

implementation which were based upon felt needs. VSO–RAISA has been guarded. Given the evidence at hand, an Action Learning cycle informed by Sen's Capability Approach has supported several deep and significant changes in the outlook and approach. The process is continuing.

Returning to the meta-issues

How can VSO–RAISA engage with and influence the values of donors, be they governments or multilateral institutions? This question represents, perhaps, the broader contribution of the Capability Approach (and other similar approaches) to the design of development interventions.

This brings us back to the concerns raised in the introduction, in which 'development' is undertaken as an ideological project. At first glance the message – that putting into practice an Action Learning process informed by Sen's Capability Approach can better support freedom – does not appear particularly radical and should appeal to donor agencies operating in democratic countries where freedom of speech is supposed to be a cornerstone of political life. However, although the message may not in itself be radical, the practical consequences would be: donors would have to give up power. Power is not something that is given up easily. The form of Action Learning process modified by Sen's approach is a fruitful topic for debate, not only within VSO–RAISA, but also with other practitioners, policy makers, and academics, for enabling challenges to power.

In order to challenge ideological projects in development and to guard the guardians, two suggestions could be explored, as follows.

The Capability Approach would be strengthened by drawing on the literature of deliberative democracy. Crocker (2006: 155–97) asserts that using this literature would help to develop procedures and principles for evaluation. James Bohman, a deliberative-democracy theorist, has already begun to examine the Capability Approach within the context of the theory of Deliberative Democracy (Crocker 2006: 167).

A second suggestion, and one which we presented earlier, is to synthesise the Capability Approach with a rights-based approach. The recent shift towards rights-based approaches by ActionAid, DFID, CARE, and in VSO–RAISA has emphasised the political nature of the development process, because power relations among the development actors have to be transformed. Sen (2005) himself has highlighted the contribution that the capability and rights-based approaches can bring to each other.

Freedom is at the core of this synthesis. However, House (2005) and Denzin and Lincoln (2005) observe that the development arena has ideologically and methodologically shifted to the right. This means that approaches to development are less likely to incorporate processes that enhance freedom, target group or beneficiary participation, and mutual accountability, where universal human rights and justice are non-negotiable. One way of challenging the power structures in the development arena is to establish evaluation and Action Learning processes and mechanisms that promote freedom. The importance of supporting freedom in evaluation frameworks is summarised powerfully by Hayek (quoted in Sen 2002: 604):

> *The importance of our being free to do a particular thing has nothing to do with the question of whether we or the majority are ever likely to make use of the possibility. It might even be said that the less likely the opportunity to make use of freedom to do a particular thing, the more precious it will be for society as a whole. The less likely the opportunity, the more serious will it be to miss it when it arises, for the experience it offers will be nearly unique.*

Action Learning cycles of development should not close off opportunities just because the implementing and donor agencies' approaches to PME have a limited informational base. How can we ensure that valued opportunities remain open? *Quis custodiet ipsos custodes?*'

Who will guard the guardians when projects are evaluated and opportunities are potentially closed off? The findings of this study show that opening the development process through implementing the freedom-based Capability Approach contributed to safeguarding freedoms that the target group value by providing opportunities for holding implementing agencies to account.

Concluding remarks

Certain monolithic power systems whose methodologies are exclusionary and biased in favour of the powerful are damaging people's abilities to realise their freedoms. This article argues for an approach akin to taking a pick-axe to these monoliths. The result may seem a mere scratch at first, but it is a good beginning. Can rights-based approaches that support freedom be more fully amalgamated within VSO–RAISA's programme design? How can the use of freedom-oriented approaches to Action Learning be more broadly explored? How can implementing agencies guard the donors? These questions, regarding the distribution of power in development, if they are addressed by reasoned debate and if they incorporate input from the disadvantaged, can challenge those in power – the guardians.

Acknowledgements

Thanks go to the regional staff of VSO–RAISA and Murray Leibbrandt, Anna McCord, and David Clark, who helped to solidify our thinking on the Capability Approach. Despite all this support, influence, and facilitation, any inadequacies in the text remain our own.

Notes

1. Although one of the authors of this paper is an employee of VSO–RAISA, the views expressed here are written in his personal capacity and do not necessarily represent those of VSO–RAISA.
2. The authors recognise the gender issues involved in development field work. Although the issue is not explored in depth in this article, exploring gender explicitly is part of the focus-group methodology: gender relations are probed during implementation. There is certainly a role for Action Learning in redressing gender imbalances. A slightly different study from the one presented here could fruitfully explore how the Capability Approach can support this objective.
3. For the distinction between cultivated truth and reality, a debt is owed to J. K. Galbraith (Galbraith 2004:1).
4. A number of more complete reviews give a broad overview of Sen's critiques, ethics, and distinctions; see, for example, Alkire (2002).
5. For a more extensive description of the methodology and an example of the operational process used by one of the authors, see Porter (2007).

References

Alkire, Sabina (2002) *Valuing Freedoms: Sen's Capability Approach and Poverty Reduction*, Oxford: Oxford University Press.

Crocker, David (2006) 'Sen and Deliberative Democracy', in Alexander Kaufman (ed.) (2006) *Capabilities Equality: Basic Issues and Problems*, New York: Routledge.

Denzin, Norman and Yvonna Lincoln (eds.) (2005) *The Sage Handbook of Qualitative Research: Third Edition*, London: Sage.

Fiedrich, Marc and Anna Jellema (2003) *Literacy, Gender and Social Agency: Adventures in Empowerment*, Tonbridge: DFID Education Publications.

Galbraith, John Kenneth (2004) *The Economics of Innocent Fraud*, London: Penguin.

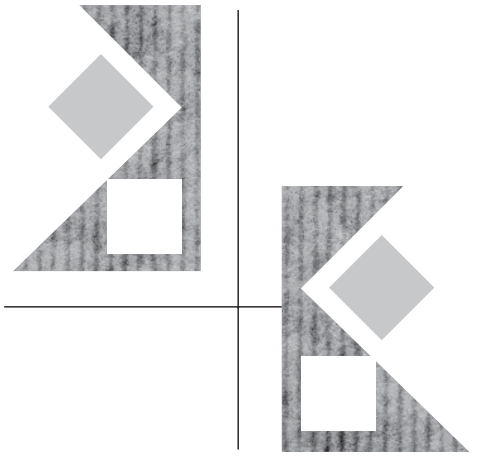
Gasper, Des (2002) 'Is Sen's Capability Approach an adequate basis for considering human development?', *Review of Political Economy* 14 (4): 435–61.
Gasper, Des (2004) *The Ethics of Development*, Edinburgh: Edinburgh University Press.
House, Ernest (2005) 'Qualitative evaluation and changing social policy', in N. Denzin and Y. Lincoln (eds.) (2005).
PEPFAR (2003) *Five Year Strategic Plan*, Washington, DC: US Government.
Porter, Stephen (2007) 'Who Will Guard the Guardians Themselves? Contributions of the Capability Approach to Capacity Development Evaluation Frameworks', unpublished Master's Thesis, Cape Town: University of Cape Town.
Sen, Amartya (1992) *Inequality Re-examined*, New York: Sage.
Sen, Amartya (1999) *Development as Freedom*, Oxford: Oxford University Press.
Sen, Amartya (2002) *Rationality and Freedom*, London: Harvard University Press.
Sen, Amartya (2005) 'Human rights and capabilities', *Journal of Human Development* 6 (2): 151–66.
US Congress (2003) 'United States Leadership Against HIV/AIDS, Tuberculosis, and Malaria Act of 2003', *Public Law 108–25*, Washington, DC: US Congress.

The authors

Stephen Porter (corresponding author) is the Monitoring and Evaluation Officer for the OVC Project for AMREF in South Africa. At the time of writing he was M & E and learning adviser with Voluntary Services Overseas–Regional AIDS Initiative of Southern Africa (VSO–RAISA). He is also a Research Associate with the Southern African Labour Development Research Unit (SALDRU), in the Department of Economics at the University of Cape Town.

Jacques de Wet lectures in development sociology in the Department of Sociology at the University of Cape Town.

Appendix B: Porter, S. (2011). ''Helping' as an Evaluation Capacity Development Strategy in South Africa'. In R. C. Rist, M-H. Boily, & F. Martin (Eds.), *Influencing change: Building Evaluation Capacity to Strengthen Governance.* Washington, DC: World Bank.



INFLUENCING CHANGE

Building Evaluation Capacity to Strengthen Governance

Ray C. Rist
Marie-Helene Boily
Frederic Martin

THE WORLD BANK
Washington, D.C.

1818 H Street NW
Washington DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org

1 2 3 4 14 13 12 11

ISBN: 978-0-8213-8403-9
eISBN: 978-0-8213-8497-8
DOI: 10.1596/978-0-8403-9

Cover design: Naylor Design, Inc.

Library of Congress Cataloging-in-Publication Data

Influencing change : building evaluation capacity to strengthen governance / Ray C. Rist, Marie-Helene Boily, Frederic Martin, editors.

p. cm.

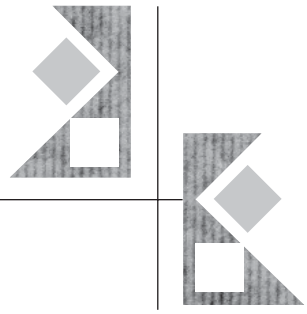
ISBN 978-0-8213-8403-9 — ISBN 978-0-8213-8497-8 (electronic)

1. Organizational change. 2. Administrative agencies—Evaluation. 3. Executive departments—Evaluation. 4. Bureaucracy—Evaluation. I. Rist, Ray C. II. Boily, Marie-Helene. III. Martin, Frédéric.

HD58.8.I5378 2011

658.4′06—dc22

2010037927



CHAPTER 9

Recognizing "Helping" as an Evaluation Capacity Development Strategy

Stephen Porter

The theory of "helping," as conceptualized by Edgar Schein (2009), provides a valuable guide to applying evaluation capacity development work. This chapter argues that when helping is applied, the results of evaluation capacity development are improved. This argument is based on reflections on implementing a community-based project monitoring system.

It is useful to think of the potential of helping in terms of the theater. Putting on a show requires a range of skills from the playwright, director, actors, and audience. In evaluation capacity development, all of these roles

The author is particularly grateful to (in alphabetical order) Addis Berhanu, Melusi Ndhlalambi, Phathisiwe Ngwenya, Ravi Ram, Patricia Rogers, and Rita Sonko for their comments on earlier drafts of this chapter, as well as to all of the field staff of AMREF who helped implement the monitoring system, especially Ntombi Mabindisa, Itumeleng Masia, Lovemore Mhuriyengwe, Juju Mlungwana, and Nonhlanhla Mthimkhulu. Thanks also go to all of the contributors to this book for their comments during the expert meeting at IDEA International

need to be played by those leading the change—the helpers. The theory of helping guides the roles of an evaluator: theory applier and model developer are equivalent to playwright and director; advocate and listener are equivalent to actors and the audience.

The beauty of Schein's theory is that it is practical. He states that the essence of relationships can be found in two components: economics and theater. These two components form the basis of a theory of helping. When enacted, the theory of helping guides evaluation capacity development to recognize and react to the relationships being built. For example, evaluators can have 10 evaluation texts on their desks. In implementing program evaluation systems, evaluation professionals may be able to use 15 percent of them. Given the law of diminishing returns, if there is a process to help project staff implement 20–30 percent of selected evaluation theory, larger returns would result from evaluation through improved implementation and use. Schein's theory of helping is one part of an effort to increase the volume of evaluation theory that can be put into practice.

The case to which helping is applied is that of the African and Medical Research Foundation's (AMREF) Bana Barona/Abantwana Bethu project, funded by the U.S. Agency for International Development/U.S. President's Emergency Plan for AIDS Relief (USAID/PEPFAR). Working in two districts in South Africa (Umkhanyakude in KwaZulu-Natal and Sekhukhune in Limpopo), the project aims to ensure that empowered children realize their rights, community-based partners operate effective childcare systems, and local municipalities implement national child policy. Both of these districts are "presidential nodes"—areas of extreme poverty in which indicators such as child and maternal mortality are worse than the national averages. In 2007, the monitoring system of the project was recognized to be in crisis. Eight months later the system was recognized as reporting reliable data. Explaining how this happened using helping as an analytical framework is the main subject of this chapter.

The chapter is structured as follows. The first section defines key terms. The second section describes helping and relates it to other areas of evaluation practice. The third section analyzes key points in the development of the monitoring systems over an 18-month period, using helping as a

in Quebec, especially Bali Andriantseheno and Mohammad Jaljouli for their insights. Finally, thanks to all staff from the community organizations, who are too numerous to mention, and the coordinators and data capturers who have stuck with the project and made monitoring happen: Zinhle Gumede, Sbongile Khumalo, Sibongile Mahalngu, Sello Makofane, Tacha Malaza, Joyce Mdluli, Sifiso Mfekayi, Dudu Mhlanga, Jabu Mlambo, Sindi Mthethwa, Zanele Mthombeni, Thembelihle Qwabe, and Phumzile Vilakazi.

reference point. The last section provides some concluding remarks and identifies areas for future work.

Definitions

In this chapter, evaluation is seen as "a key analytical process in all disciplined and practical endeavors" (Scriven 1991, 1). This means that evaluation is applicable to a range of activities from products to programs to personnel and beyond. Scriven (1991, 1) defines evaluation as a "process of determining the merit, worth and value of things." The definition of the Organisation for Economic Co-operation and Development (OECD) for good program evaluation (box 9.1), in alignment with Patton (2008), essentially asks "What? So what? Now what?" The question of "now what?" adds an extra dimension of utilization beyond Scriven's definition. Within this chapter, the OECD definition of evaluation is used because of this extra component. Monitoring is seen as a subset of evaluation. It is a different form of evaluation from, for example, impact evaluation. It helps answer different questions about a program.

A definition of helping, described in the next section, is "a basic relationship that moves things forward" (Schein 2009, ix). This definition covers a wide array of help, which can be placed on a continuum from formal to informal (Schein 2009). Informal help covers giving directions and behaving appropriately toward others (using good manners). Semiformal help involves payment and less personal involvement for some kind of service (purchasing a piece of equipment or providing assistance in using software). Formal help involves formal agreements and the provision of professional expertise (employing a lawyer, doctor, or consultant). In this

Box 9.1 The OECD Definition of Evaluation

According to the OECD, "an evaluation is an assessment, as systematic and objective as possible, of an ongoing or completed project, program, or policy; its design; implementation; and results. The aim is to determine the relevance and fulfillment of objectives, developmental efficiency, effectiveness, impact, and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors."

Source: OECD 1991, 5.

chapter, the main types of help analyzed are semi-informal and formal help, where different roles and forms of inquiry become more pertinent than in informal help.

The Components of Helping

For Schein (2009, 29), the essence of relationships is contained in economics and theater. The "implication for would-be helpers is to become conscious of social economics and the social theater that we all live in, to think clearly about the helper role … and to assess what sort of currency and what kinds of values must be managed to make the relationship fair and equitable." This brief quotation sums up the links between helping and evaluation capacity development. The two components are interactive—the economic exchange is defined by the theatrical roles various players take on. For good evaluation capacity development, the helper needs to know what the demand is for evaluation, what values are to be measured, and how the relationships in the evaluation are to be managed.

In this section, the interaction of economics and theater are expanded upon and related to evaluation practice. The core principles of helping are summarized in box 9.2.

An alternative to helping is understanding the task only in terms of outputs. In practice, sometimes there is pressure to get things done. Success in developing evaluation capacity, for example, may be measured in the number of workshop participants or the budget spent. As Wiesner points out

Box 9.2 The Core Principles of Helping

Principle 1. Effective help occurs when both giver and receiver are ready.

Principle 2. Effective help occurs when the helping relationship is perceived to be equitable.

Principle 3. Effective help occurs when the helper is in the proper helping role.

Principle 4. Everything you say or do is an intervention that determines the future of the relationship.

Principle 5. Effective helping starts with pure inquiry.

Principle 6. It is the client who owns the problem.

Principle 7. You never have all the answers.

Source: Schein 2009.

in chapter 2, the effectiveness of evaluation is defined by how the demand is structured for improved results and performance. When short-term goals are the driving force, the direction is inappropriate and rewards are perverse; independent, credible, and useful evaluation capacity cannot be developed. Training and other short-term interventions are valid when the right conditions are established; when they are the sole measures of success, the process can become distorted. One way to establish the right conditions is to work through the social economics and theater involved in defining an intervention (Schein 2009).

The basis of exchange is interactions between people. When economic systems are built on trust and a sense that the exchanges are fair and equitable, confidence and efficiency within a system are supported (Mertens 2009). This basis of exchange can be seen as demand and supply; each side needs the other. The same is true with evaluation capacity development. On the demand side, a decision has been made that evaluation capacity needs to be developed and help has been sought. On the supply side are people who claim to be able to help strengthen evaluation capacity. To some extent, a social process underlies this exchange in which intangibles exist between the supply and demand. Appreciation of the intangibles in a relationship sometimes entails slowing down and seeking understanding rather than pushing to get things done. Schein (2009) recognizes that there is a power imbalance in the relationship between helper and helped that can intrinsically affect the success of an intervention and the changes (outcomes) that can be realized.[1]

Schein (2009) outlines 11 possible pitfalls in establishing a helping relationship, 5 for the client and 6 for the helper. These pitfalls result from genuine anxieties, inequalities, and ambiguities arising in exchange. For example, there can be resentment and defensiveness on the part of those being helped. These attitudes may be expressed as the withdrawal of some participants from a workshop at the last minute. The people supplying evaluation capacities (henceforth called *helpers*) are one up on those demanding it. They are the experts, upon whom, to some extent, the client is dependent. In this formal role of experts, helpers can dispense their wisdom prematurely (Schein 2009). They can uncover too much, shaming current efforts. A better approach is to be an appraiser, treating the program and the staff with respect in a process of dialogue. Schein (2009) moves beyond merely describing the social economics by defining and providing guidance on how to move between different helping roles to work as an appraiser rather than a bully.

Schein (2009, 48) points out that "at the beginning of any helping situation the appropriate roles and the rules of equity are inherently

ambiguous . . . both the helper and the client have to develop an identity and choose a part to play." For undertaking evaluation capacity development, it is helpful to think of four roles. Schein (2009) delineates three helping roles: expert, doctor, and process consultant. The liberty is taken here to expand this to four roles to more closely match evaluation capacity development work by breaking out the role of process consultant into audience and actor. The four roles are as follows:

- Playwright: the expert evaluator producing conceptual frames and documents
- Director: the person who moves people through an evaluation system and deploys the tools of evaluation
- Actor: the person playing the evaluation role with the client, working through and demonstrating how things work in practice
- Audience: People watching others conduct the evaluation, giving appropriate praise through cheers and applause.

The helper plays all four of these roles, sometimes in the same day. Acting out tasks with the client leads to reflection; being the playwright means updating the documents. The helper can then return to the audience and watch others play out tasks, then become the director and try out new props and routines. The helper does not take on the problem but facilitates change. The client owns the problem.

It is in the selection of roles that the success of evaluation capacity development is defined. If the client is misread, the work of the helper will be misdirected. Supply will not meet demand. To help reduce the likelihood of miscommunication, Schein (2009, 66) defines *humble inquiry* as "the key to building and maintaining the helping relationship"; approaching evaluation capacity development using humble inquiry equips the helper to enter dynamic situations in "a supportive, ego enhancing way." Schein recommends starting out in a process consultant role as the most effective way to establish fairness and help uncover the real demand for help. In the above schema, this is equivalent to starting off as the audience before moving to humbly working through the current system with the client to create a climate for deeper understanding and trust in which both parties reveal more of themselves (Schein 2009).

Schein augments these roles by defining four forms of inquiry (table 9.1). It is this detail on roles and forms of inquiry that separates Schein from a number of other authors in development and evaluation work. Many authors indicate that having an interactive relationship is a good thing, but they leave it to the practitioner to muddle through how to do so.[2]

Table 9.1 Schein's Four Forms of Inquiry

Type of inquiry	Purpose	Roles	Sample questions
Pure	Build confidence and status of the client. Develop context to reveal anxiety, feelings, and information. Diagnose issues and plan for action.	Audience and actor. Evaluator watches issues play out and at times performs tasks with client (for example, checking data entry together for errors). This passive but attentive role should be balanced by "constructive opportunism," in which significant elements are revealed that enable another form of inquiry.	Tell me more. . . . When did this last happen? Can you give me some examples?
Diagnostic	Influence client's mental processes by focusing on issues other than the ones the client chose to report, in terms of feelings and reactions, causes and motives, actions taken or completed, and systemic questions.	Audience and actor. Evaluator watches issues play out and at times questions the work of the client (for example, asking why certain data errors keep occurring). This is a passive role, but the evaluator starts to be a more influential actor with the client.	How did you feel about that? (Feeling and reaction) Why do you think you are having this problem? (Cause and motive) What have you tried to do so far? (Action taken) How will your colleagues react? (Systemic question)
Confrontational	Articulate analysis by making suggestions and offering options.	Director and playwright. Actors are directed to take new positions. The playwright may write up and work on new processes.	Did that make you angry? Could you do the following?
Process oriented	Focus on interactions between client and helper to make client conscious of the helper's influence. This can be combined with the other forms of inquiry.	Director. Works through issues with actors, enabling examination of the relationship between the client and the helper.	Are we getting anywhere? Are my questions helping you?

Source: Author.

Recognizing the importance of social economics and theater in these four forms of inquiry provides an accessible and usable approach to developing relations for evaluation capacity development. Helping encompasses developing horizontal relationship between the helper and the helped rather than vertical teacher and student relations. It requires being aware of

the small things that take place in the relationship between consultant and client. It guides moments when the helper senses that those being helped recoil because they are close to a change. Helping occurs when helpers become learners, recognizing the limits of their knowledge in a given context. Helping is about being humble about the limits of formal education. It is certainly not micromanagement, as there is give and take in working toward a shared direction. Sharing a journey is a central notion in helping, with a particular emphasis on the challenge of monitoring changes in someone's behavior within a change process.

Helping is largely about utilizing knowledge of social economics and theater in establishing human relationships to incrementally learn together by focusing on the interpersonal responses. Social economics are the expectations on either side of the exchange; theater mediates how the exchange happens. In working from this perspective, helping resonates with other writings on evaluation and development. In chapter 2, Wiesner highlights the importance of the demand for improved results. Boyle and Lemaire (1999) emphasize the importance of the location and structure of evaluation demand and supply. Toulemonde (1999) outlines a framework for thinking about the interaction between demand and supply for evaluation, mixed in with a little theater in describing the use of carrots, sticks, and sermons. Patton's (2008) description of situational responsiveness for evaluation is in many ways a description of how best to match the demand for evaluation while playing different roles. To enable this, Patton takes on the role of the active-reactive-interactive-adaptive evaluator. Chambers (2007) seeks to locate evaluative demand through participatory processes, employing theater through various participatory rural appraisal techniques.

In summary, helping is similar to and compatible with a large variety of research and evaluation practice in the way it describes developing relationships for change. Schein's theory of helping, though not as detailed as the work of Kusek and Rist (2004), Fetterman and Wandersman (2005), Gustavsen (2006), Senge and Scharmer (2006), or Mertens (2009), presents an accessible and usable set of procedures that can be applied to guide practice. The applicability of helping is its major advantage and the reason why it is applied as a prism through which to view the following case study.

Applying Helping to Understand an Evaluation of the Bana Barona/Abantwana Bethu Project

This section examines the experience of developing monitoring capacity in eight grassroots organizations in South Africa. Within a period of eight

months, the system moved from being in crisis to being recognized by the donor (USAID/PEPFAR) as in good practice.

Often for things to change and for people to seek help, a crisis is required. This section analyses the implementation of the monitoring system using helping as a prism through which to view how progress was made. It suggests that when helping processes are followed, relationships are built, technical evaluation tools are easier to embed in project design, and results improve. This reflection is based on the experience of the author, who was the monitoring and evaluation (M&E) officer responsible for designing and implementing the system, and reflections from staff, partners, and consultants.

Phase 1: Pure Inquiry and Initial Attempts at Change

The project monitoring system was not started from scratch. Technically, the system was of a reasonable standard. A detailed M&E plan was in place, data-collection forms were being filled in, and a project database had been designed. Despite this, the project had no good idea of what its reach was, where key bottlenecks lay, or how well the project was being implemented. The project had just been reviewed, and data quality was found to be substandard. The system was suffering from a lack of regular help, given that no M&E officer had been in place for six months. The tools that had been developed were not embedded and were not evolving.

Guided by the project staff, the M&E officer tried to get an overview of the system across all sites, through site visits to community partners, review of documents, and interactions with the grant managers of this project, Pact. The outcome of this pure inquiry was the development of some supportive tools and the holding of a training workshop on these tools.

In terms of applying the theory of helping, three points stand out for this phase. First, the initiation of helping followed a process of pure inquiry, with the roles of audience and actor being assumed. From watching how the forms were processed, it was found that the manual counting of service delivery forms, the main method for ascertaining outreach, was a bottleneck. Manual counting was complicated by the need to differentiate between different levels of servicing (a child receiving three or more services was counted differently from a child receiving fewer than three) and different categories of services within a population of about 5,000 children. Doing this meant creating complicated tally charts and lists between which children moved as months progressed. Watching this process, asking questions, and meeting with Pact clarified the demand for evaluation, which came principally from the implementing agency, AMREF, but also from

the community organizations, which wanted to reduce the manual counting effort and report their work accurately. AMREF's demand for evaluation stemmed from the need to get a better overview of the project and to report with confidence. This initial period, during which pure inquiry was undertaken, served as a solid foundation for moving forward. Relationships started to be built, system blockages understood, demand for evaluation located, and questions developed.

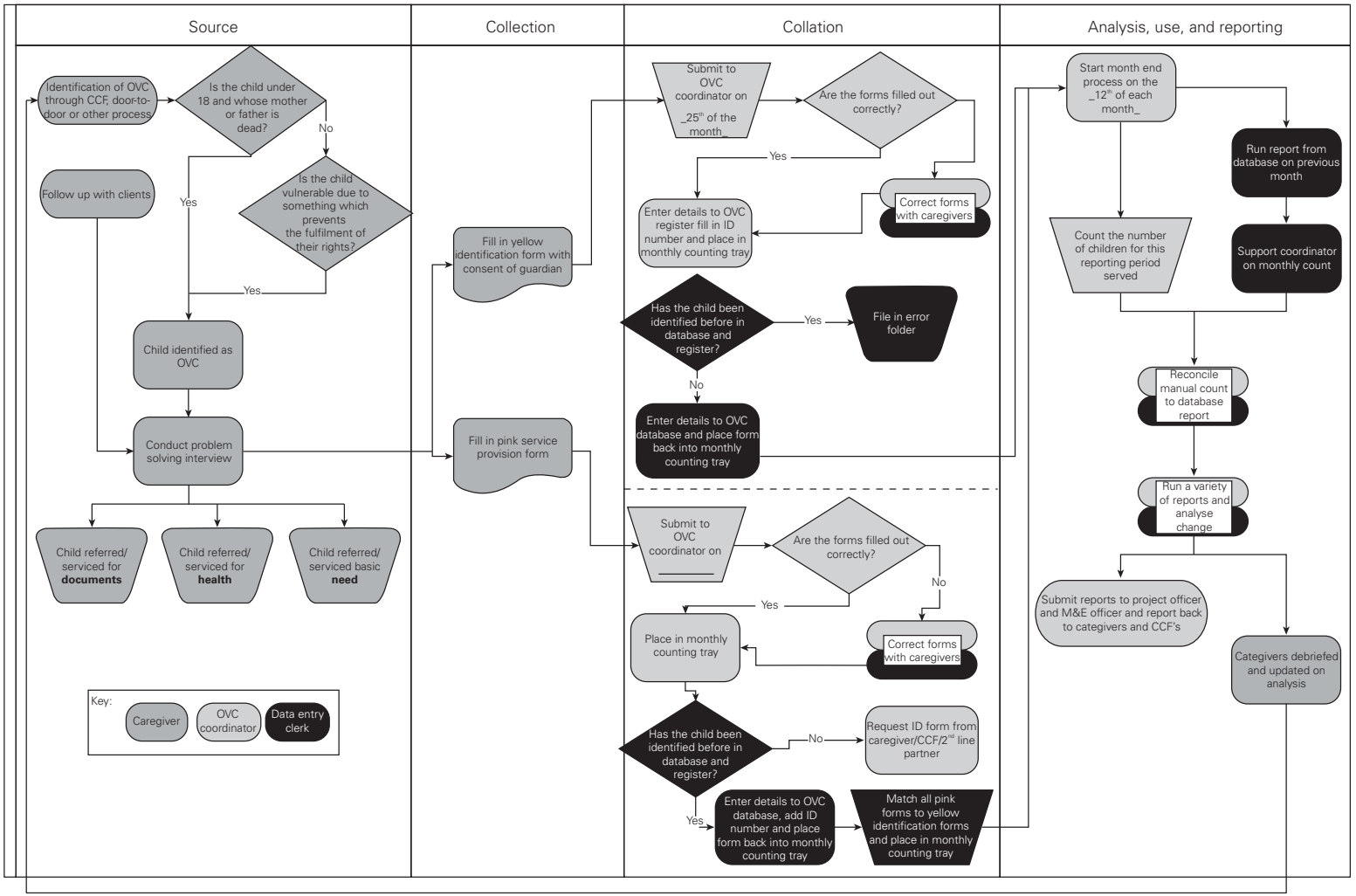
Second, technical evaluation skills helped frame the capacity development response. The evaluator rapidly moved between the roles of playwright, audience, and actor. In addition to listening, there was also a need to go back to the office to try to make sense of what had been heard. Two tools were valuable in this regard: the Barefoot Collective's (2009) way of thinking about organizational development and the development of a detailed data flow (figure 9.1) based on written and technical support from Pact (McCoy and others 2008).

Third, recognizing the broad phase of organizational development helped tailor the helping assumptions on which the systems development was based. The community organizations being partnered with are in a pioneering phase of organizational development. This phase is characterized by flexibility of approach, very few policies or procedures, a great deal of experimentation, and little planning (Barefoot Collective 2009). Given this new monitoring, processes can be introduced through consultation with a small group of people who work closely with the bottlenecks and are clear on why they demand change. This approach can be contrasted with a more rational type of organization, which is defined by clear leadership, professionalism, plans, policy, and systems (figure 9.2).[3] Within this type of organization, change can be more difficult, because negotiation is with higher-level personnel who "own" the system and may not be directly exposed to the bottlenecks and the demand for change (Barefoot Collective 2009). This knowledge helped direct the first conversations with the organizations and define who needed to buy into the system.

The data-flow process mapped the actions and paper flow that takes place from the identification of a child through the analysis and reporting of the entire project. Mapping the flow of data enabled a single language to be spoken about the bottlenecks in the process. From this mapping, a number of issues could be raised, in more diagnostic-style inquiry, within the project.

The use of these tools demonstrates that technical tools from practical evaluation texts need to underscore the helping process. Initial conversations and the building of relationships will not develop into useful evaluation capacity if, at some stage, the team does not take on the role of playwright.

Figure 9.1 Project Data Flow for AMREF Project



Source: AMREF 2008.

Figure 9.2 The Barefoot Collective's Phases of Organizational Development

Source: Barefoot Collective 2009.

What is necessary is not to be dominated by this form of approach but rather to use it once the process of pure inquiry has been undertaken in association with diagnostic inquiry.

Leading from these two processes—pure inquiry and initial redesign—a workshop was organized to discuss the analysis and support the implementation of the redesigned elements of the system. The workshop covered some basic definitions of M&E and worked through some of the new tools, such as the project data flow. The outcomes of the workshop were mixed. The data flow was taken up as a shared language of the monitoring system. The new manual counting procedures did not work, however. Looking back, it may have been a little early to move from a pure to a more directive, even confrontational, form of inquiry in relation to the counting procedures. The bottleneck was identified, but the context in which it operated was not fully understood. The problem came to light six weeks later, when the next phase of reporting was due and uncertainty over the numbers being reported remained.

In summary, evaluation capacity development workshops need to be tailored to meet issues in context. Doing so requires that the issues first be understood through pure inquiry and then translated into technical evaluation tools.

Phase II: Success after Learning from Others

In the second phase of developing the monitoring system, there was a marked shift in emphasis. More on-site support was provided in which issues were worked through with individuals and organizations. Inspiring

this shift of approach was learning from the implementation of a computer database for tracking and monitoring, the Soweto Care System, which was being rolled out across AMREF's community partners at roughly the same time as the manual counting was being completed.[4] The database was a substantial improvement over the previous one. It is a well-designed, off-the-shelf system that reduces the work needed for manual counting and includes easy-to-use backup procedures, allowing the quality of the data to be assessed centrally. As feedback from the field stated, "the database is . . . crucial to us, to know how many children we're servicing at our finger tips. . . . the Soweto Care System was a lifesaver for us, because time was saved from manual counts."

It was not the tool itself but the way the system was rolled out that was all important for the development of evaluation capacity. The implementation of this database followed helping practice. Based on their experience from some 50 organizations, the consultants identified demand as the most important factor for successfully implementing the database. Demand for good data had already been established within AMREF and community partners. During the initial identification of issues in the existing monitoring system, the pure inquiry uncovered some issues regarding the previous database. A diagnostic inquiry process was then undertaken to develop a new database system that addressed these issues. This process culminated in program staff working jointly through the key fields of the Soweto Care System, which was then rolled out. In short, a number of the tools that can be associated with helping were used in the preliminary phases of implementing the system. Following the design workshop, the consultants moved to training at the site level. Training involved data capturers and at least one other member of staff. The consultants took on a director's role, directing others to do the acting, never "touching the keyboard" themselves.

Ongoing support was given by AMREF staff to embed the Soweto Care System and to implement other processes within the data-flow process (see figure 9.1). Managing the implementation of the system using the data flow took about 70 percent of the M&E officer's time for four months, with 50 percent spent on site. The on-site support process used the full range of inquiries described in table 9.1. Looking back at the work done on site, the diagnostic form was used about 50 percent of the time, pure inquiry was used about 30 percent of the time, and confrontational and process-oriented inquiry each took about 10 percent of the time. This support focused on spotting issues in data quality, advising on approaches to remedy issues with data collection, identifying gaps in filing, and verifying issues with the database. Some of the interactions in this process are

shown in table 9.2. The confrontational and process inquiry approaches are valuable when used sparingly.

Evaluation capacity development is a change process, which, as Machiavelli pointed out in *The Prince*, can be difficult, doubtful, and dangerous. This means that at times there will be resistance to new behaviors. Sometimes an extra push is required to get through difficult issues. The process of evaluation capacity development in the Bana Barona/Abantwana Bethu project did not follow the helping methodology exactly. Mistakes were made and apologies proffered. At some times, helpers were viewed with joy; at others, they were viewed with suspicion and even anger. Given these reactions, helping can be emotionally draining.

At the operational level, seven of the eight community organizations that AMREF works with were generally committed to the process at any one time. The uncommitted organization changed: when one organization's commitment started to wane, extra effort was put in.

Table 9.2 Examples of On-Site Support Linked to Helping Roles and Inquiry

Interaction	Example	Type of inquiry	Role
Problem solving together	Working through a reconciliation of a database report to the forms. This was an incremental process that helped both parties understand issues in data entry.	Pure and diagnostic	Audience and actor
Laughing about issues	Developing personal relationships with the data capturers based on trust; seeing the humorous side of mistakes (for example, not taking double entries in the database too seriously)	Pure and diagnostic	Audience and actor
Taking issues raised about the database seriously	Recording issues identified by data capturers and escalating them to developers	Diagnostic	Director
Creating some competition between data capturers	Talking about how many entries different data capturers had achieved and using that metric as a yardstick to push people when commitment waned	Confrontational	Director
Pushing through on issues when resistance was met on challenges the community organization's staff could resolve	Repeating exercises of analysis in areas that were difficult, such as data analysis	Confrontational and process oriented	Director

Source: Author.

Help was given mainly to community partner organizations rather than AMREF internal staff. For AMREF staff, this was an unacceptable gap in their own capacity development. The demand for improved monitoring was immediate, as monitoring data had become a critical project issue. Attention was therefore focused at the source of data, the community. As a result, AMREF staff were not helped as extensively during this period.

The initial focus on the community had two unintended longer-term effects. First, information use is still patchy. Some organizations use it solely for reporting, whereas others are starting preliminary analysis of care worker caseloads. Second, because the help was focused at the operational level, the managers of community organizations became less involved. Had local staff been more fully involved, it is possible that they could have better supported information use while getting more buy-in from managers. A theory of helping needs to be applied at different levels of the organizations involved, even when they are at the pioneering phase. Different levels of demand for evaluation have to be taken into account. This speaks to the issues that Heider raises in chapter 5 with respect to establishing an enabling environment for evaluation.

Outweighing these issues are the outcomes. First, the quality audit revealed an immediate improvement in the data. Such an audit is an important process in judging the quality of the monitoring system for USAID/PEPFAR. Data quality is measured in the areas of validity, reliability, integrity, precision, timeliness, and completeness. At the head office level, precision rose from 23 percent to 100 percent, and reliability rose from 79 percent to more than 96 percent; at the community level, validity rose from 79 percent to 92 percent, and precision rose from 59 percent to 87 percent. In the short term, these improvements satisfied the demand for evaluation at the AMREF level. The evaluation team was able to enter into detailed discussions with partners about the way the project operates, how their staff work, and how the management support decisions—all based on up-to-date, reliable evidence.

Second, through the analysis of information, the project team refined its ideas, leading to a new results framework and innovative new projects. The emphasis is now on the importance of the coordination of care by different role players. The monitoring system aided the development of this focus area. Monitoring the project allowed the evaluators to analyze where referrals were breaking down between the community and the health system. In response, community-friendly tools were developed that are being used to support the coordination of care. One such tool is a poster, which was constructed through a process of pure and diagnostic inquiry as monitoring data became available. The poster shows services available within a South

African municipality. This poster is used in a number of ways. For example, care workers use it to discuss with children and guardians the kinds of rights they have. The poster relates to a referral list and could be used in conjunction with a list of telephone numbers of health service providers in local area.

In responding to the monitoring data, mobile technology is being explored for the purposes of developing community health information systems. Through better understanding of the challenges care workers face in working with health services, mobile technology has been developed to help smooth communication and information exchange. The partnerships developed over 12 months reached the stage at which mobile technology is now being piloted.[5]

This technology includes an area-based directory (a toll-free number that provides a directory of area-based service); a health announcements system, which provides SMS airtime to care workers to ask questions and interact with clinics and other service providers, such as schools; and a service rating system, in which clients text a number to rate their satisfaction with a service such as a clinic or police station. These changes demonstrate that although the helping roles of director, actor, and audience require intense communications, the playwright function of those developing evaluation capacity cannot be forgotten. Quiet periods are required to integrate technical evaluation theory into practice to move systems forward.

Coming out of the data analysis processes was recognition of the need to further support changes in quality at the source of information. Service providers are likely to have left formal education early. This means that training processes, forms, and data-collection methods need to be tailored to work with their knowledge base. A training process for implementing service protocols was implemented that drew on the lessons of the past; it was grounded in a solid diagnostic process, mentoring, follow-up, and the involvement of AMREF staff. Interestingly, the training for care workers integrated some of the main tenets of helping. Open-ended questioning and relationship-building techniques were integrated into the training. AMREF staff, care workers, and coordinators report that the training process changed the way they operate with clients, reinforcing the value of recognizing helping as an evaluation capacity development strategy.

This analysis shows that where helping was used, albeit unconsciously, successes in evaluation capacity development occurred. A number of other interventions, such as peer and external support to data capturers and coordinators, were undertaken during this phase to support the development of evaluation capacity. Because of space constraints, it is not possible to

describe them here. However, where they were most successful, roles and inquiry were undertaken in the mode suggested by Schein (2009).

Phase III: Ongoing Work

Following from these successes, a third phase is now in process. It can be seen as ongoing development, where the realities of staff turnover and changing context become relevant. Working through a process of developing evaluation capacity does not have an obvious ending point. Systems are in flux, with mixed results. Perhaps the best that can be hoped for is the self-reorganization and regeneration of monitoring systems to higher and lower levels of functioning.

Country-level systems are now being implemented using results-based management (Kusek and Rist 2004) as a guiding frame. Although this is an organization-wide process, demand also comes from within the projects, which want to more systematically collate and analyze their experiences. Within this process, helping can assist with smoothing the recognized implementation and management issues related to results-based management systems (Perrin 1998; Kusek and Rist 2004). Helping is of use because it is specific about the roles that need to be played. For example, the role of actor and audience can be undertaken in the development of outcomes and indicators. Meanwhile, the director can reinforce the demand for the evaluation capacity. The playwright can work behind the scenes, translating the steps within a monitoring plan and framework.

This new phase also needs to be cognizant of some of the limitations of the helping approach, which requires ample resources (people, finances, and time). In addition to the full-time M&E officer, there was the Pact M&E adviser, the AMREF corporate M&E leader, and external advice from RMIT University in Melbourne. About $70,000 (3.6 percent of the budget) was expended over 12 of the 18 months. This money supported one full-time salary for eight months, data capturers, traditional workshops, additional expertise, the rollout of the Soweto Care System, and travel to the site for mentoring. Helping also takes a large amount of time in the field—potentially 40–60 percent, depending on the level of intensity. Given these resource requirements, there is a need to learn from the previous phases and be efficient in implementation.

New challenges exist in using helping in the ongoing development of monitoring systems. Relationships need to be defined, new staff oriented, and resources used effectively and efficiently. In this ongoing work, the roles of audience, actor, director, and playwright will still need to be performed by

those developing evaluation capacity. This process will be more conscious of the theory of helping.

Conclusion

Using helping supported the integration of technical evaluation approaches to capacity development work through improved personal relationships. As the outcomes suggest, when the importance of social economics and social theater roles are recognized, even unconsciously, energy is released and new pathways for change opened.

Schein's (2009) theory of helping is useful in developing evaluation capacity. Helping does not replace other strategies and technical approaches; it complements them by assisting them in becoming operational in a given context.

Schein goes beyond other theorists in the accessible way in which he describes how to go about developing evaluation capacity. Schein is accessible to the organizational practitioner. He provides guidance on where to be opportunistic, charismatic, and systematic in a way that can complement other approaches.[6]

Useful work could be undertaken to further evaluate helping. Results for evaluation capacity development can be judged against the standards contained in three OECD evaluation principles: independence, credibility, and utility of evaluations (OECD 1991; see chapter 5 of this book). Using these principles to judge the effectiveness of helping could help go beyond this introductory case study by using a standard evaluative frame for capacity development strategies. In the longer term, doing so will help develop fuller knowledge about the strengths and limits of this approach and others.

One year after the end of the period of intense support, the monitoring system continues to be embedded and is to some extent self-regulating. Many of the gains were retained during a period of program staff changes, although progress was hampered. The robustness of the system and the use of helping are related. This connection will continue to be used in the ongoing development of monitoring systems within AMREF.

Notes

1. The recognition of imbalances in power resembles arguments in Akerlof (1970), where, because of differences in information, incentives exist to supply a lower standard of good. Although this issue is not explored in this chapter, it is an area that requires further exploration.

2. Mertens (2009) is less detailed in her guidance. Bawden (2006) is more academic, describing the "epistemic transformation" of the evaluator.
3. Two additional phases are described in Barefoot Collective (2009): the integrated and associative. These two are not described here, because they are not directly relevant to the case.
4. The database was developed by a volunteer placed with a community organization by Volunteer Services Overseas (VSO) in partnership with VX Company in the Netherlands as part of a corporate social responsibility initiative.
5. Two main partners are involved in the mobile technology project, Cell Life, based in Cape Town, and HIV911, based in Durban.
6. See chapter 5 of this book, on ordered chaos.

Bibliography

Akerlof, G. A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84 (3): 488–500.

AMREF (African Medical and Research Foundation). 2008. *How We Care: The Monitoring and Evaluation Plan September 2008–September 2010*. Nairobi: AMREF.

Barefoot Collective. 2009. *The Barefoot Guide to Working with Organisations and Social Change*. Cape Town: Barefoot Collective.

Bawden, R. 2007. "A Systemic Evaluation of an Agricultural Development: A Focus on the Worldview Challenge." In *Systems Concepts in Evaluation: An Expert Anthology*, ed. B. Williams and I. Iman, 35–46. Point Reyes, CA: Edge Press.

Boyle, R., and D. Lemaire, eds. 1999. *Building Effective Evaluation Capacity: Lessons from Practice*. Piscataway, NJ: Transaction Publishers.

Boyle, R., D. Lemaire, and R. C. Rist. 1999. "Introduction: Building Evaluation Capacity." In *Building Effective Evaluation Capacity: Lessons from Practice*, ed. R. Boyle and D. Lemaire, 1–19. Piscataway, NJ: Transaction Publishers.

Chambers, R. 2007. "Poverty Research: Methodologies, Mindsets and Multidimensionality." Working Paper 293, Institute of Development Studies, Brighton, United Kingdom.

Fetterman, D. M., and A. Wandersman. 2005. *Empowerment Evaluation Principles in Practice*. New York: Guilford Press.

Gustavsen, B. 2006. "Theory and Practice: The Mediating Discourse." In *Handbook of Action Research: Concise Paperback Edition*, ed. P. Reason and H. Bradbury, 17–26. Thousand Oaks, CA: Sage.

Kusek, J. Z., and R. C. Rist. 2004. *Ten Steps to a Results-Based Monitoring and Evaluation System: A Handbook for Development Practitioners.* Washington, DC: World Bank.

McCoy, L., N. Njeri, E. E. Krumpe, and R. Sonko. 2008. *Building Monitoring, Evaluation and Reporting Systems for HIV/AIDS Programs.* Washington, DC: Pact.

Mertens, D. M. 2009. *Transformative Research and Evaluation*. New York: Guilford Press, OECD/DAC.

OECD/DAC (Organisation for Economic Co-operation and Development/Development Assistance Committee). 1991. "DAC Principles for Evaluation of Development Assistance." Paris.

Patton, M. Q. 2008. *Utilization-Focused Evaluation*. Thousand Oaks, CA: Sage.

Perrin, B. 1998. "Effective Use and Misuse of Performance Measurement." *American Journal of Evaluation* 19 (3): 367–79.

Schein, E. H. 2009. *Helping: How to Offer, Give, and Receive Help*. San Francisco: Berrett-Koehler.

Scriven, M. 1991. *Evaluation Thesaurus*. Thousand Oaks, CA: Sage.

Senge, M., and C. O. Scharmer. 2006. "Community Action Research: Learning as a Community of Practitioners, Consultants and Researchers." In *Handbook of Action Research: Concise Paperback Edition*, ed. P. Reason and H. Bradbury, 195–206. Thousand Oaks, CA: Sage.

Toulemonde, J. 1999. "Incentives, Constraints and Culture-Building as Instruments for the Development of Evaluation Demand." In *Building Effective Evaluation Capacity*, ed. R. Boyle and D. Lemaire, 153–75. New Brunswick, NJ: Transaction Publishers.

Appendix C: Porter, S., & Goldman, I. (2013). 'A Growing Demand for Monitoring and Evaluation in Africa'. *African Evaluation Journal, 1*(1), 9 pages. Available from: https://doi.org/10.4102/aej.v1i1.25

This article has built on some of the first comparative work undertaken across Africa on M&E systems, building not from a donor demand, but in-country demand to share experience (the project was a partnership between DPME and CLEAR). As a result, there are emerging pan-African efforts to build on each other's experience, with an active exchange programme happening between South Africa, Benin and Uganda around evaluation. International organisations like CLEAR, 3ie and donors are actively supporting this sharing. This should help to reduce donor dominance, both in terms of concepts and instruments, help to reinforce in-country capacity to develop M&E systems, and build local confidence.

Further in-depth work is needed, which will both help to deepen the analysis and also lead to more in-depth sharing across countries. Some fruitful follow-up work could be undertaken from four perspectives: (1) citizens, (2) line ministries, (3) parliaments and (4) the profession of evaluation. In this analysis, there is a gap in knowledge of how citizen demands for development spur government demands for evaluation. Filling this gap would be important given the appearance of an increasingly active citizenry on the continent. An investigation of line ministries would give a deeper political perspective on how the centralised rules and incentives play out in practice. The use of M&E information by parliaments provides an opportunity for increased demand and use of M&E information for accountability. Parliaments are locations of latent demand for evaluation, where there is space for contestation around evidence. Finally, deeper analysis of the profession of evaluation would give an indication of the gaps between government demand and the current supply as governments start to regulate the markets they generate as they commission evaluation.

Acknowledgements

The authors would like to acknowledge the inputs of Salim Latib and Anne McLennan for valuable comments on earlier versions of this article, as well as the two blind reviewers. Funding for the cases underlying this article was received from GIZ and the Centers for Learning on Evaluation and Results trust fund housed by the World Bank.

Competing interests

The authors declare that they have no financial or personal relationship(s) that may have inappropriately influenced them in writing this article.

Authors' contributions

S.P. (Centre for Learning on Evaluation and Results for Anglophone Africa and University of the Witwatersrand) was lead author and co-manager of the case research. I.G. (Department of Performance Monitoring and Evaluation of the Presidency of South Africa) was chief sponsor in the DPME of the case research, supporting development, writing and editing of the article.

References

AfrEA, 2007, 'Making evaluation our own: Strengthening the foundations for Africa-rooted and Africa led M&E: Summary of a special conference stream and recommendations to AfrEA', in A. Abandoh-Sam et al. (eds.), *Evaluate development, Develop Evaluation: A Pathway to Africa's future*, Author, Niamey.

Barber, M., 2008, *Instruction to Deliver: Fighting to Transform Britain's Public Services*, Methuen Pub. Ltd., London.

Behn, R.D., 2003, 'Why measure performance? Different purposes require different measures', *Public Administration Review 63*(5), 586–606.

Bemelmans-Videc, M.L., Rist, R.C. & Vedung, E., 2003, *Carrots, sticks, and sermons: Policy instruments and their evaluation*, Transaction, New Brunswick.

Benington, J. & Moore, M.H., 2011, 'Public value in complex and changing times', in J. Benington & M.H. Moore (eds.), *Public Value: Theory and Practice*, Palgrave Macmillan, London.

Boyle, R. & Lemarie, D., 1999, *Building Effective Evaluation Capacity: Lessons From Practice*, Transaction Publishers, London.

Chelimsky, E., 2006, 'The purpose of evaluation in a democratic society', in I. Shaw, J.C. Green & M.M. Mark (eds.), *The Sage Handbook of Evaluation*, pp. 33–55, Sage, Thousand Oaks, CA.

Chen, H., 2005, *Practical Program Evaluation: Assessing and Improving Planning, Implementation and Effectiveness*, Sage, Newbury Park, CA.

Centre for Learning on Evaluation and Results (CLEAR) & The Presidency Republic of South Africa Department: Performance Monitoring and Evaluation (DPME), 2012a, *African Monitoring and Evaluation Systems: Exploratory Case Studies*, University of the Witwatersrand, Johannesburg.

CLEAR & DPME, 2012b, *African Monitoring and Evaluation Systems: Workshop Report*, University of the Witwatersrand, Johannesburg.

DPME, 2011, *National Evaluation Policy Framework*, Author, Pretoria.

Engela, R. & Ajam, T., 2010, 'Implementing a Government-wide Monitoring and Evaluation System in South Africa', in Independent Evaluation Group (ed.), *Evaluation Capacity Development*, Working Paper Series 21, World Bank, Washington.

Funnell, S.C. & Rogers, P.J., 2011, *Purposeful Program Theory: Effective use of Theories of change and Logic Models*, Jossey-Bass, San Francisco.

Lopes, C. & Theisohn, T., 2003, *Ownership, Leadership and Transformation: Can We Do Better for Capacity Development?*, Earthscan/James & James, New York.

Mackay, K., 2007, *How to Build M&E Systems to Better Support Government*, World Bank, Washington.

Ministry of Finance and Economic Planning (MoFEP), 2011, *The Budget Statement and Economic Policy of the Government Of Ghana for the 2012 Financial Year*, Author, Accra.

National Development and Planning Commission (NDPC), 2011, *Resources spent on M&E and Statistics Final Report*, Author, Accra.

NDPC, 2012, *National Development and Planning Commission Staff*, viewed 28 August 2012 from http://www.ndpc.gov.gh/DG's Office.html%3E

National Treasury, 2012, *Estimates of National Expenditure*, Author, Pretoria.

Organisation for Economic Co-operation and Development (OECD), 2005, *Paris Declaration on Aid Effectiveness and the Accra Agenda for Action*, Author, Paris.

Owen, J.M., 2007, *Program Evaluation: Forms and Approaches*, 3rd edn., Guilford Press, New York.

Patton, M.Q., 2008, *Utilization-focused evaluation*, 4th edn., SAGE, London.

Perrin, B., 1998, 'Effective Use and Misuse of Performance Measurement', *American Journal of Evaluation 19*(3), 367–79.

Picciotto, R., 1995, 'Introduction: Evaluation and Development', *New Directions for Evaluation 1995*(67), 13–23.

Picciotto, R., 2012, 'Country-Led M&E Systems -Robert Picciotto' *Part 2*, video, Youtube, viewed 20 April 2012 from http://www.youtube.com/watch?v=UfNakPTODBs%3E.

Plaatjies, D. & Porter, S., 2011, 'Delivering on the Promise of Performance Monitoring and Evaluation', in D. Plaatjies (ed.), *The Future Inheritance: Building State Capacity in Democratic South Africa*, pp. 292-312, Jacana, Johannesburg.

Pollitt, C., Bouckaert, G. & Van Dooren, W., 2009, *Measuring Government Activity*, OECD, Paris.

Ramkolowan, Y. & Stern, M., 2009, *The Developmental Effectiveness Of Untied Aid: Evaluation Of The Implementation Of The Paris Declaration And Of The 2001 DAC Recommendation On Untying ODA To The LDCS*, Development Network Africa, Pretoria.

Toulemonde, J., 1999, 'Incentives, Constraints and Culutre-building as Instruments for the Development of Evaluation Demand', in R. Boyle & D. Lemaire (eds.), *Building Effective Evaluation Capacity: Lessons from Practice*, pp. 153–176, Transaction Publishers, New Brunswick.

Vedung, E., 2003, 'Policy instruments: typologies and theories', in M.L. Bemelmans-Videc, R.C. Rist & E. Vedung (eds.), *Carrots, sticks, and sermons: Policy instruments and their evaluation*, pp. 21–59, Transaction, New Brunswick.

Wiesner, E., 2011, 'The Evaluation of Macroeconomic Institutional Arrangements in Latin America', in R.C. Rist, M. Boily & F. Martin (eds.), *Influencing change: evaluation and capacity building driving good practice in development and evaluation*, pp. 23–40, World Bank, Washington.

Appendix D: Porter, S. (2013). 'A Change of road for a rights-based approach? A case study of mobile technology in KwaZulu Natal, South Africa'. In P. Gready & W. Vandenhole. (Eds.), *Human rights and development in the new Millennium: Towards a theory of change*. London: Routledge.

12

A Change of Road for the Rights-Based Approach?

A reflection on piloting a health-enabling mobile technology programme in KwaZulu-Natal, South Africa[1]

Stephen Porter

Introduction

In order for human rights-based theories of change to be relevant for grassroots development organizations, they need to incorporate conceptions of rights that tangibly direct practitioners to programme actions. This argument is based on a reflection of the experience of implementing a mobile technology pilot, called Impilo, which sought to improve people's access to health services in Umkhanyakude in the north of KwaZulu-Natal, South Africa. A tangible conception of rights directs attention to processes that help to apply key concepts of human rights in programming. The current UN human rights-based development framework is too wide-ranging, and is difficult to apply at the grassroots of development work. The principle of joint production of information between public service producers and users is posited here as a potential avenue to sharpen rights-based theories of change. This principle has relevance to managing partnerships and employing multidimensional strategies to enhance human rights where mutual accountability is problematic or where organizations are mobilizing the public, while also undertaking collaboration and joint work with the government (see chapter 1 in this volume).

A theory of change explains how an initiative will be put into action, and how it will cause a desired change. A theory of change should have two components: an action theory and a change theory (Chen, 2005; Funnell and Rogers, 2011).

A programme can be well implemented, but not cause the desired change because it is targeting the wrong mechanism. Likewise, a programme can be implemented in a way that is different from the original design, but still achieves interesting results.

A greater understanding of the limitations of the current rights-based approach was gained as Impilo (meaning health/life in isiZulu) moved through a number of phases of development. The Impilo project started in 2008 and sought to improve people's ability to claim their right to health through attempting to improve their interaction with health services based on access to improved information. In attempting to utilize a theory of change based on the UN rights-based framework, difficulties arose with the scope of the demands for grassroots development work. Consequently, the concepts used in Impilo were limited. In working through this process, the potential role of joint production as a useful framing principle became increasingly prominent as a way in which to implement programmes that improve rights-based claims.

If integrated into a programme, joint production can inform both the action and change models. In practice, joint production takes the form of service providers (duty-bearers) and people (rights-holders) regularly entering into discussion and occasionally redefining action around different interpretations of data. Working in this way supports the use of knowledge in decision-making by enhancing the salience and legitimacy of the data being produced (Clark *et al.*, 2006). The formulation of joint production draws principally on work conducted by the Packard Foundation (2010), but resonates with work across academic boundaries. Joint production has a solid foundation of empirical literature as a mechanism to improve institutional interface with the public. To date, however, joint production as a critical mechanism to rights-based theories of change has not been conceptualized explicitly.

This chapter tells the story of Impilo's attempts to integrate a rights-based approach. It does this by, first, providing an explanation as to why the rights-based approach should be important in the formulation of Impilo. Second, the important connection between rights-based claims and changes in information is elaborated. Third, the reasons for the selection of the UN rights-based approach to guide the theory of change are reviewed. Fourth, the main elements of UN rights-based theories of change are described. Fifth, placement of rights-orientated concepts within the Impilo programme is analysed. Sixth, issues with current rights-based theories of change are discussed. Seventh, the potential of the joint production principle is analysed. Finally, it is concluded that for the rights-based approach to become relevant it needs to embark on a different road that centralizes improved information flows in its theories of change.

Why the rights-based approach should be important

in the formulation of Impilo

Research shows that in South Africa there is a gap in accountability between the state and its citizens. A recent World Bank (2010: 3) report argued that in South Africa 'political accountability, communication with all citizens must be improved and institutionalized, particularly among the poor and those living in informal settlements'. The disconnect between people and state, which has been accompanied by rapidly emerging class structures, has been argued to be an important factor in the breakdown of the delivery of government services in South Africa, and has had widespread consequences (World Bank, 2010: 2–3; SWOP, 2011).

Rights-based approaches for a development programme should provide a powerful tool set to enhance the interface between people and state. People's interactions with governments are an important space in which to place demands for improved delivery of government programmes. If people are able to assert their demands upon a government that is receptive, then services can be better attuned to the needs of the population by pointing services downwards towards desired capabilities rather than upwards to bureaucratic rules. Within development, human rights is seen as a tool of political transformation, as a moral imperative, and as a measurement tool for development. Nyamu-Musembi and Cornwall (2004: 47–48) argue that implementing a rights-based approach properly means that there can be 'positive transformation of power relations among the various development actors', for which 'the potential gains are of significance beyond the arena of human rights, signalling possibilities for the transformation in the very ways in which we think about "development"'. Essentially, a rights-based approach should provide a set of tools to support the public in securing accountability from the state.

There are visible accountability disconnects between public and state in the area where Impilo was piloted: Umkhanyakude in northern KwaZulu-Natal. People in Umkhanyakude live in the second most deprived district in South Africa. In terms of the deprivation index, a composite index of education, household, water, electricity and unemployment data compiled by StatsSA, Umkhanyakude is ranked 51 out of the 52 districts. In health it has the second lowest TB cure rates and highest incidence of sexually transmitted infections (Day *et al.*, 2010). Access to health services in Umkhanyakude is poor. People need to travel long distances to reach care facilities and then wait in long queues. The roads are mostly dirt, while tarred roads, built mainly for military purposes during apartheid, are poorly maintained and removed from where people live. In addition, a persistent problem noted during the implementation of Impilo was people not knowing where their closest health services were located. These issues compound waiting times, as people come for services on days when they are not provided.

Anecdotal evidence collected during site visits over a period of two years indicates that although individual staff mem-

bers can be responsive, beyond the frontline institutional accountability becomes weaker. A story of hospital equipment helps to illustrate this point. Over a period of two years, some hospital equipment awaiting collection and transport to the local hospital was left outside. It was once reasonably usable, but slowly disintegrated over time. Where beds once were, now rusty health hazards stood in a yard. The community organization that witnessed this would not challenge the hospital for fear of repercussions. The local clinic attempted to intervene but the hospital, to which the clinic reported, did not respond. The rights-based approach should provide accessible tools to give people the confidence to make claims on services and improve their interactions with the health system to bridge such accountability gaps. Dormant hospital equipment can be a first point of interaction.

However, pockets of good healthcare practice do exist in Umkhanyakude. Perinatal mortality has in general been on a downward trend since 2004 (Day *et al.*, 2010). Child immunization and follow-up antenatal care seem to be in reasonable shape in the district (KwaZulu Natal Department of Health, 2010). Health expenditure in this district is some of the highest per capita in South Africa, with a large number of staff per client. It should also be remembered that South Africa is a democracy, where rights have been enshrined in a constitution. This provides the basis for rights-based claims to be made. In this context, a rights-based approach should provide guidance to build on areas where rights are realized and define actions to reduce accountability deficits and improve the lingering health issues.

Linking rights-based claims to changes in information

The concept of a rights-based claim is of central importance for those seeking to articulate a rights-based theory of change for development programmes. Human rights can seemingly be achieved when a measurable desired change occurs. For example, a large reduction in child mortality can be achieved without any transformation in power relations between the people and state. In this context, as Donnelly (2003: 9) argues, 'rights would remain not only out of sight but out of mind as well'; actual realization of rights requires that people have the power to claim their rights, as this 'distinguishes having a right from being the (rights-less) beneficiary of someone else's obligations'.

The prioritization of the claim as a core component in a theory of change for human rights is reinforced by a number of scholars. Donnelly (2003: 12) argues that 'Human rights *claims* express not merely aspirations, suggestions, requests, or laudable ideas, but rights-based demands for change' (emphasis added). This argument links to Nyamu-Musembi and Cornwall's assertion that power relations can be transformed by a rights-based approach. If people do not express a claim, there is no focal point for challenging a system of power. For Amartya Sen, the process of making a claim is central to the

achievement of freedom. In discussing the importance of rights, Sen consistently highlights the importance of the process of deciphering, publicly reasoning and asserting claims to human rights (Sen, 2009: 360, 62, 65–66, 85–87).

Making claims is central to a rights-based theory of change. Improving people's ability to claim their rights in a development programme means supporting people to mobilize and utilize local capacity to put pressure on their government to provide progressively for the realization of their rights. This is as opposed to developing a programme where outside aid agencies replace the role of government in the provision of services, an approach that can last only as long as the external agency does. In Umkhanyakude, this would mean that instead of usurping the health system through direct provision of services, a rights-based approach would direct agencies to find ways to mobilize people and work with government, while drawing on existing knowledge and experience in community development. The role of the rights-based approach then becomes to help people to analyse and develop projects that respond to the underlying issues.

In a context where the state is still open to engaging with people around accountability issues, a concrete basis is required for this engagement. The concept of the claim needs to be made operational for programming. The production and dissemination of information between people and the providers of services is a useful entry point. Power cannot be redistributed unless there is a tangible base of information that backs up a claim with evidence. Enhancing flows of information is problematic, however. Incentives exist for informational asymmetry.

George Akerlof's (1970) classic article 'The market for "lemons"' highlights the importance of informational asymmetry in the quality of the delivery of goods and services. He uses the example of the automobile market to demonstrate that, where the seller of a used car has more information than the buyer, incentives are produced to up the price and conceal defects in the car. In the same way, where an area of government (it is important not to see government as monolithic) has more information on the provision of services, incentives are produced to offer lowerquality services. Looking for problems in information flow can help to understand where accountability problems exist, while simultaneously providing a mechanism for change. Access to information is important for strengthening people's claims. As Stiglitz (2002: 461) argues:

> There are asymmetries of information between those governing and those governed, and just as markets strive to overcome asymmetries of information, we need to look for ways by which the scope for asymmetries of information in political processes can be limited and their consequences mitigated.

For example, internationally it has been recognized that improved accountability through communication and information exchange between the state and citizens in middle-income countries can lead to improved health outcomes (Bloom *et al.*, 2008).

A rights-based theory of change should enable information to be used to improve how people claim rights and how

government responds to claims. Having and using information to reinforce a claim is an important part of the sustainability of changes in rights. Applying information to a problem increases the power of people, especially the least well off, as demonstrated by ActionAid's reflect methodology (Duffy *et al.*, 2009: 3). Remedying information asymmetries helps to direct us to mechanisms that can be used to strengthen claims in Umkhanyakude. A good rights-based theory of change should adequately conceptualize how to improve the flow of information from government to citizens and the subsequent use of information.

Selection of UN rights-based approach as guiding framework at the inception of Impilo

The project team of Impilo took as its starting point United Nations (UN) documentation that explains the rights-based approach. At the inception of Impilo it was seen as important to be able to make arguments on the basis of legal entitlements. Legally grounded arguments should provide a degree of strength to the arguments made for Impilo. The rights-based approach is defined by the Office of the High Commissioner on Human Rights (OHCHR, 2006: 15) as:

> a conceptual framework for the process of human development that is normatively based on international human rights standards and operationally directed to promoting and protecting human rights.

According to this definition, an organization implementing a rights-based approach would draw upon the local and international legal bases to support people to make claims upon government. The UN documentation was initially selected to guide the implementation of Impilo because it details and elaborates a development-oriented approach to rights connected to the international legal framework.

Although the South African constitution does provide for a range of socioeconomic rights, how these would be implemented in development programmes is not clear. It was hoped that the UN approach would provide practical guidance that would be supportive of implementation. In retrospect, the pertinence of some of the programme approaches developed in South Africa by the Treatment Action Campaign (TAC) probably provides a better starting point for developing Impilo. However, the link between the TAC's court action and its rights-based approach to HIV treatment is not well covered in the literature. The legal battle to win access to antiretroviral medication is emphasized in texts (e.g. OHCHR, 2006: 3). Yet the importance of their community-level treatment literacy campaign as a rights-based programme

approach received limited coverage. This campaign achieved widespread mobilization of poor people through information dissemination and developing people's ability to interact with the state (Heywood, 2009).

Though alternative conceptual bases are available for developing a rights-based theory of change, owing to limited capacity in terms of money, time and staff, a range of literature and techniques could not be worked through. At the start of the programme the UN literature was most accessible and detailed. A closer examination of the UN rights-based approach framework and two methods that seek to put it into operation highlights the useful, and the impracticable, elements for development organizations.

UN rights-based theories of change

A theory of change describes how an initiative will be put into action, and how it will cause a desired change. A theory of change should have two components: an action theory and a change theory (Chen, 2005; Funnell and Rogers, 2011). The action model details the protocols, people and relationships that will make a project work. The change model details the mechanisms that need to be activated in order for a change to come about. Both are essential. The term 'programme theory' incorporates a number of different approaches to defining how a programme works, such as a logical framework or intervention logic (Funnell and Rogers, 2011).

The UN rights-based theory of change is grounded on the Stamford Declaration (UNDP, 2003). This declaration outlines a range of principles and processes for implementation. The third area of common understanding, which is the most important part of this document for the development of a rights-based programme theory, states that:

> Development cooperation contributes to the development of the capacities of 'duty-bearers' to meet their obligations and/or of 'rights-holders' to claim their rights.

Conceptually, this third area of common understanding centralizes the interface of people and government around rights-based claims. Its refinement can be seen as the keystone in a UN rights-based theory of change (and indeed Impilo). This is because it has consequences for both the action and change models of a programme. It directs attention to the development of capacity for people to demand, and the government to respect, protect and fulfil rights (action theory). It is expected that the resulting changes in capacity would improve the ability of the parties to engage around the rights-based claim (change theory).

The work of Jonsson and Madsen, who align their work to the Stamford Declaration, is drawn upon here to explore further this third area of common understanding. Madsen's work has a more strongly defined change model; Jonsson's has a more fully defined theory of action. Both place the development of capacities to claim rights as central to putting into operation frameworks grounded in a UN rights-based approach (Jonsson, 2005: 51; Madsen, 2007: 126).

Madsen's change model, represented in Figure 12.1, places the interface around rights-based claims at the heart of the programme theory. Although Jonsson does not use this model explicitly, many of the same concepts and ideas are used in explaining the importance of the interface between rights-holders and duty-bearers (Jonsson, 2005: 56).

In terms of an action model, Jonsson's approach is more developed than Madsen's. Jonsson's (2005: 55) action model focuses on improving five components of capacity, so that:

- duty-bearers understand problems that should be addressed
- people know they can take action
- people develop strategies to access and control resources
- people have the ability to access and use communication tools
- both rights-holders and duty-bearers undertake evidence-based assessment and decision-making.

In summary, taken together, Jonsson and Madsen provide an approach to rights-based programmes that moves beyond narrower legal understandings of the realization of rights – what Madsen (2003, cited in Madsen, 2007: 13) calls a 'Rule of Law paradigm'. They provide a framework to which the Impilo programme theory initially aspired. This theory of change focuses closely on the third point of common understanding within the Stamford Declaration. The change model recognizes and specifies the need to improve the interface capacity between rights-holders and duty. The action model outlines an approach to developing capacity. However, the utility of this approach in actual grassroots development is problematic. The range and scope of demands on capacity is too great in the resource-constrained setting where implementation was attempted. Furthermore, with the exception of Jonsson's demand for an improved evidence base for decision-making, the conceptualization of the role of information to strengthen the claim remains largely absent. This is explored in the following section.

Placement of rights-orientated concepts within the Impilo programme

In October 2008 the Impilo Mobile Technology pilot was first defined during interactions between the author of this chapter, then a monitoring and evaluation officer of the African Medical and Research Foundation (AMREF), and the director and programme staff of HIV911, an organization specialising in collecting and disseminating information on formal and informal health providers across South Africa. In implementing programmes to mitigate the impact of HIV and AIDS, AMREF identified problems in the interface of the local health system. Health providers did not know the range of health services being provided in their area. There were particularly strong disconnects between formal health providers (clinics and hospitals) and informal health provision (community-based organizations). This meant that community care workers, who often provide the link between the formal and informal system, did not have complete information on a range of issues, such as where to refer people; when the most appropriate time to access services was; what documentation was required; and what level of service to demand. These issues point directly to a problem in the accountability and interface capacity in the health system. In this way Madsen's theory of change, presented in Figure 12.1, becomes relevant because there appears to be a lack of engagement on how to make changes to the process of making claims.

From examining these issues, assets within both organizations were identified. AMREF had a community-based programme in the North of KwaZulu-Natal that worked across various levels of the health system, both formal and informal. HIV911 had a data set of formal and informal health providers that it had been looking to disseminate more widely, as well as links to mobile technology companies. Having recognized the near-universal access to cell phones, both organizations also wanted to explore innovative ways of disseminating and tracking health information through mobile technology. The monitoring and evaluation officer at AMREF had also been looking at how to incorporate rights-based concepts into the evaluation systems, so had some understanding of human rights concepts.

Initial project documentation in late 2008 envisaged four interfaces that drew upon mobile technology; in 2010 a fifth was added. Each of these interfaces and the

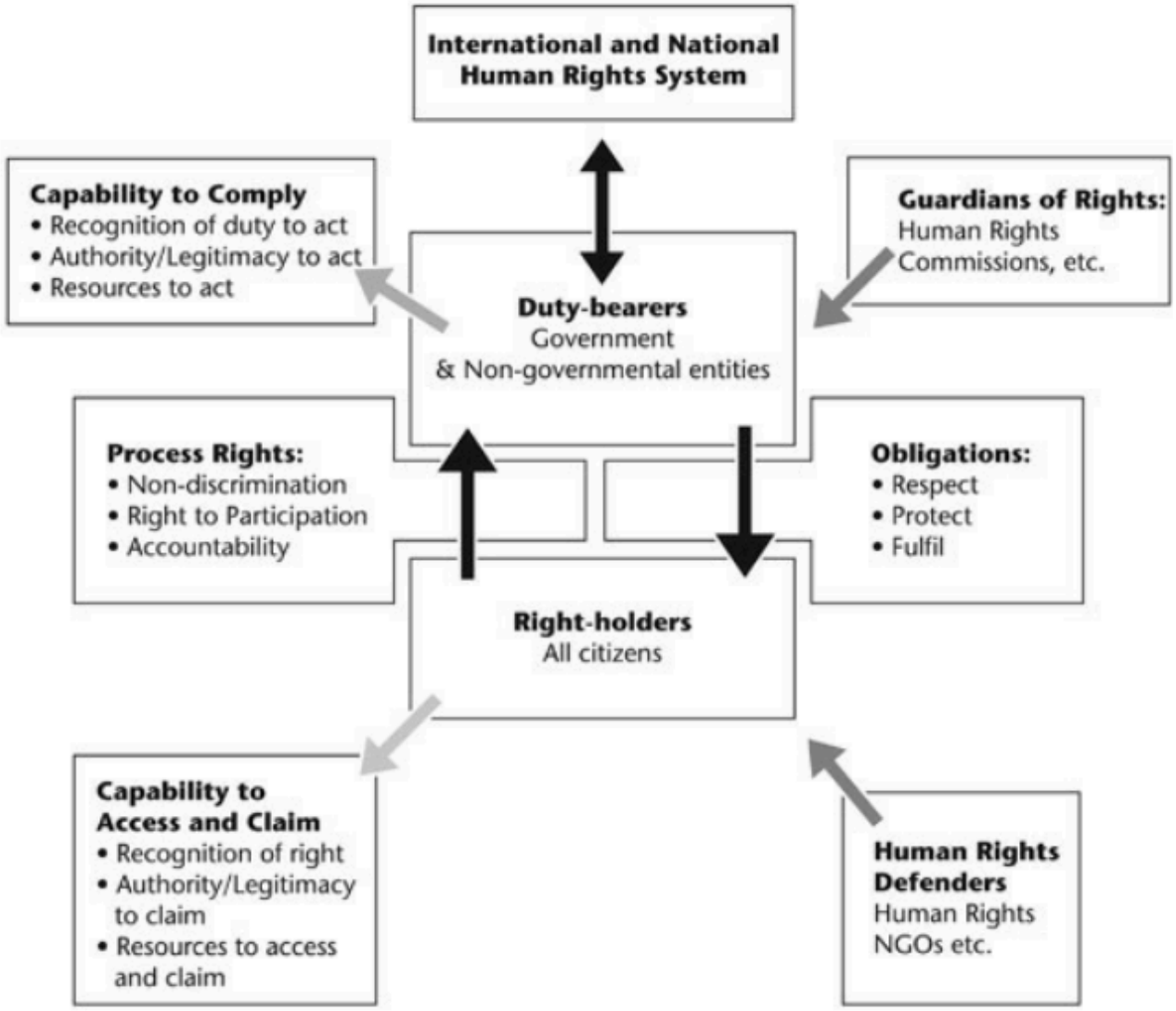


FIGURE 12.1 Madsen's RBA Navigator (Madsen, 2007: 128)

processes to implement them can be related to Jonsson's theory of action, which emphasizes capacity development to improve rights-based claims, as demonstrated below.

The first, and most successful, mobile technology interface enables people to find local health providers. After using a

menu system on their phone, they were sent a text message (SMS) giving information on three of the closest organizations that gave the services they required. This system was rolled out following a Department of Health radio campaign on HIV counselling and testing, linked to World AIDS day in 2010. Over a three-month period the system received 1685 individual requests for information across South Africa. Of these, 1056 were for counselling and testing-related services, 173 were for social grants, and 128 were for rape-survivor support. This demonstrates the demand for improved sources of information directly from the population.

The second mobile interface sought to provide answers to frequently asked questions. For example, an issue identified consistently across organizations in Umkhanyakude was that people did not have information on how to obtain the documents required to access services. In this instance it was hoped that people would use a mobile system to find local channels to access the documentation that would enable people to get better social assistance. This system was partially available, though not to the extent originally sought, at the time of writing this chapter. These first two mobile interface tools promote Jonsson's fourth area of capacity development by enabling people to access and use communication tools so that they have improved capacity to find out more about services.

Third, it was envisaged that a system would support referrals between selected formal health services and the informal health supporters (care workers). The care worker in this original conceptualization was a focal point for health knowledge in the community, and would announce and facilitate communication between various health providers. The original metaphor for this focal person was a star, an *Inkanyezi* in isiZulu. This star would be able to make appointments at the local clinics through a mobile interface and would get information from the clinic on specific health issues. This would include, for example, the days when the doctors were due to be in the clinic. At the time of writing, this system had not been implemented as it requires a large amount of coordination and communication from local health providers and mechanisms to incentivize the *Inkanyezi*. This third interface could potentially have developed Jonsson's first and third areas of capacity: the *Inkanyezi* could have supported strategies to better access and control resources as well as generating a shared understanding of the problems that should be addressed within the health system.

Fourth, it was noted that potential existed for generating data on local health needs. At an early stage of project development it was noted that interactions through mobile technology created bits of data. Data would be generated by the process of people demanding different types of service. This information would be automatically stored in databases. This means that by providing information on services, the mobile phone becomes a data collection tool. The potential here was for information to be accessed locally and analysed to improve health system responses. This system has been trialled with some success, demonstrating that demand for information on services closely correlates with other data sets. For example, the 128/1685 requests for information on rape-survivor support closely matches national crime data. Taking these results and feeding them back into appropriate forums potentially could develop capacity in the fifth area that Jons-

son posits, by supporting relevant evidence-based assessment and decision-making on health issues.

Finally, a new rating system was developed in the latter stages of the pilot programme to enable people to rate the formal and informal health services. This was to be trialled by some of the community organizations involved. The response to this mini-survey showed that people could use the system. People generally were happy with the service received at clinics, although there were issues with waiting times and the trust people had in the nurses at the clinics. Although this is not an absolute indicator of health system performance, it does provide some client perspectives. In addition, the speed with which information could be accessed and represented graphically (overnight if necessary) was seen as a real bonus of the system. This system could help to develop people's and service providers' capacity to understand problems by applying an evidence base to decision-making, the first and fifth of Jonsson's capacities.

This short description illustrates that, from the outset of the project, central rights-orientated concepts were being integrated into practice. The change model improved the interface around claims between people and service providers through different interactions with information. Meanwhile, the action model being implemented closely aligned to Jonsson's rights-based capacities.

Project documentation also reflected the rights-based orientation of the programme. In 2009 the rights-based language was explicit in project documents. For example, the aim of the project in the middle of 2009 was to increase access to health through assimilating health-enabling mobile phone elements into a community collaboration project. The interpretation of the right to health and the analysis of health problems was based on the definition of the right to health contained within General Comment 14 (CESCR, 2000), which is explicit on the social determinants of health, but framed rather legalistically.

Rights-based models were integrated into both the action and change theories of Impilo. As stated above, the design of the programme aspired to be linked closely to the UN documentation of the rights-based approach. It should be conceded that some of the above analysis has been applied retrospectively. Programme designers became familiar with Madsen's and Jonsson's work only in the course of 2009. However, in implementing Impilo some challenges with the UN framework arose.

Issues with current rights-based theories of change

In implementing Impilo over a period of two years, some lessons emerge that raise issues with UN oriented rights-based theories of change. The implementation of Impilo took place in a highly resource-constrained environment. Little budget

was available, while people implementing the project undertook development in addition to their normal work. This does mean that it was difficult to explore the full intricacies of the rights-based theory of change. In piloting a project or an approach, a lack of resources is not unusual, and arguably provides a stern testing ground for the real world. If concepts are not intuitive or cannot be communicated, then development practitioners put them to one side. Four main issues arose in implementing a theory of change influenced by the UN framework for human rights, specifically the work of Madsen and Jonsson.

There is an issue, first, with the guidance for developing capacity in a rights-based programme. In Impilo, both people and service providers had indicated an enthusiasm to own and use the mobile technology interfaces. This was expressed in a number of focus groups, individual interviews and responses to presentations over a period of a year (2009–10). Owing to resource limitations, local ownership and mobilization of Impilo were aimed at a few local organizations. However, these organizations did not take up Impilo as vigorously as they had indicated they would. This points to a need to think differently about capacity development for rights-based projects. Capacity development for a project such as Impilo has to take place on numerous levels. Capacity needs to be developed throughout organizations, and not just in relation to rights-based capacities. Support to project management processes and the ongoing provision of resources were required before Impilo could engage fully with developing rights-orientated capacities. For Impilo to work, organizations did not just need to see the Impilo tool set as a good idea, but required the necessary competencies to carry it forward. Although capacity development is relevant to a rights-based theory of change, a different entry point is required to focus attention on actions that enable capacity development throughout organizations.

The second issue that was identified with the UN framework and Madsen's approach is the lack of definition of strategies to iron out informational asymmetries. Madsen does identify human rights monitoring as a strategy, but this is an expert observer's approach. UN legal documents do highlight the importance of information: for example, in the International Covenant on Civil and Political Rights (ICCPR), article 19 relates to transparency. However, this is not elaborated upon when it comes to rights-based programmes. For example, in its guidance on the rights-based approach, the OHCHR (2006: 13) does cite information as important to human rights processes, for example in increasing accountability for budget transparency. However, neither Madsen nor the OHCHR fully elaborates on nor centralizes the importance of generating locally relevant information for strengthening rights-based claims. Information is there, but is not recognized as a core element in improving the claim. Jonsson (2005: 57) does recognize the importance of capacity for rational decision-making based on evidence-based assessments and analysis. Yet this argument still needs development in Jonsson's work. Gathering data, feeding back information and supporting the use of information with a population is an important process in facilitating access to rights.

Third, the scope of analysis demanded by Madsen and Jonsson is quite extensive. There is an expectation that

programmes will be able to analyse the rights-holder and duty-bearer interface, define contextualized strategies for capacity development, and then measure their contribution to a change in the realization of rights. This is to happen while checking if the overall principles of a rights-based approach are being applied as per the Stamford Declaration. For smaller organizations and NGOs, this kind of analysis is out of reach. For example, although the definition of the right to health was used to validate the approach, the full scope of analysis was never followed through. The scope of General Comment 14 is potentially infinite. Smaller organisations find it difficult to work through the appropriate use of such a wide-ranging tool as General Comment 14. Currently, neither Madsen nor Jonsson presents many options for narrowing analysis.

Evaluating the impact of Impilo from a rights-based perspective is also challenging, given the analytical scope and skills required. In demonstrating the realization of rights, one needs to specify which ones are to be targeted. To do this, legal documentation has to be drawn upon, which requires specialist skills. Claiming an impact on the realization of rights could be misleading. The short-term realization of rights can be noted: someone attaining a better service in a clinic for one day is fairly achievable. The longer-term measurement of changes in secure access to rights is not within the remit of smaller NGOs. In demanding these kinds of analysis, the UN approach calls for a technical legal understanding of rights and expansive measurement of the human rights context. This can cause difficulties in both implementation and measurement for smaller organizations.

The final cluster of issues with implementing a UN rights framework relates to the language and concepts. In working at community level, the terms 'rights-holders', 'duty-bearers', 'empowerment' and 'participation' pose challenges, as they have little relevance. These terms were not applied in the work of Impilo. On occasion they were discussed, but they have little field reality. The extent to which the use of legalistic rights-based language is valuable to programme design is questionable. Rights-based language was used to communicate and validate the Impilo project to potential donors and management, however. Legal language serves as an advocacy tool, but is not particularly helpful for programming. It is telling that the rights-based principles from the Stamford Declaration were not included in project documents. Concepts such as 'respect', 'protect' and 'fulfil' as well as 'rights-holder' and 'duty-bearer' do not appear. Arguably, such terms as 'rights-holders' and 'duty-bearers' can be seen as residual language derived from the more legalistic approach to rights.

Furthermore, even words in more regular use in development practice, such as 'participation' and 'empowerment', are not necessarily very useful. This is because they refer to shapeless processes sometimes disconnected from the ends of the project. For example, this author has attended so-called empowering participatory events that were little more than bribing people with food in order to extract opinions for reports. The UN rights-based terms lose their appeal and relevance, partly due to technical complexity and partly due to debasement.

Developing a shared understanding in order to localize legal concepts would require in-depth discussion. The work of the TAC shows us that a large amount of focus, commitment and capacity is required to make rights-based discourse

practical (Heywood, 2009). In reviewing video and documentary evidence from the Impilo pilot, people know that different forms of pressure can be brought to bear on government service providers. Yet the language used by people does not directly articulate the UN or legal rights-based concepts. A large amount of effort has to be put into vernacularizing rights-based language. This is because it is difficult to translate the concepts of UN documents into isiZulu. In resource-constrained settings, it would be better to have more immediately available concepts that direct actions to tangible change.

In summary, previous work on the definition of rights-based theories of change served as an important base for Impilo. The UN framework and the work of Jonsson and Madsen highlight important parameters for a theory of change. For example, the centralization of the claim provides a core focus for rights-based programming. Jonsson's (2005) components of capacity are especially useful in framing a programme's right-based action model. However, four conceptual issues have been highlighted. The first three refer to issues in implementation (analysis, language, and process of capacity development), while the fourth (the role of information) notes an area that could be expanded. What this highlights is the need for a framing principle that would serve as a tangible entry point to rights-based programme design. Arguably, joint production has potential as a framing principle.

Joint production

Applying the joint production principle to a programme seeks to improve the way information is translated into social knowledge, which leads to action. Joint production is envisioned here as processes where service providers (government) and people regularly enter into discussion around programme issues and occasionally redefine action. Joint production operates at multiple levels of a programme's design. In the design of a programme, it supports processes of defining data that are to be collected. During implementation, joint production refers to the social process of different interest groups coming to a shared view of what information means and then deciding to take action on the issues that emerge. The understanding of joint production in this chapter draws upon, and moves slightly beyond, the formulation applied by the Packard Foundation (2010). The Packard Foundation applies joint production mainly to the process of generating scientific knowledge that is linked to action. Here the value of joint production is perceived as an over-arching principle of programme design and operation to improve information flows and use between the public and service providers. The organization supporting joint production is then positioned between the government and public, helping to form partnerships and support mutual accountability.

The main contention of the joint production principle is that knowledge becomes useful, and used, when it meets three

tests, which are described here in descending order of importance (Packard Foundation, 2010). First, information needs to be salient. Salience can be achieved by seeking answers to important issues in time for decision-making: Does the information gathered answer weighty questions? Second, the process of understanding issues needs to be seen as legitimate by interest groups in being unbiased and inclusive. Strengthening legitimacy can be attained through the involvement of the stakeholders who would use information: Do the interests involved see the information as fair? Third, information needs to be credible. Credibility depends on the information passing appropriate tests of validation. The extent of validation depends on the issues that need to be understood and the timeline for decisions, rather than a predefined gold standard: Are the results seen as trustworthy?

Evidence shows that the application of these three tests improves the use of information. Empirically, the value of joint production is demonstrated in adaptive management for environmental conservation literature (Lee, 1999; Clark *et al.*, 2006). The empirical claims of the value of joint production are reinforced by relationships to similar concepts across academic fields. Related concepts can be identified in the public value literature, where the concept of co-production of government services is argued to play an important role in improving the accountability of government (Coats and Passmore, 2008: 37–42; Benington and Moore, 2011). In the development field, links can be made to the Reflect approach, championed by ActionAid. Reflect fuses 'the theoretical thinking of Paulo Freire with the practical visualization methodologies developed with Participatory Rural Appraisal' (Duffy *et al.*, 2009: 5). Over a period of 16 years' implementation, consistent evidence has emerged of Reflect's impact on literacy and supporting mobilization around rights (*ibid.*: 7). The contribution of the Packard Foundation's (2010) formulation is that it can bring together these multiple areas of practice under a concise and accessible set of principles for improving information use.

Application of the principle of joint production helps to work through three of the issues with current rights-based theories of change noted above (information, scope and language). The fourth issue, which relates to the targeting of capacity development, requires deeper analysis, and falls beyond the scope of this chapter. Capacity development for improving the realization of rights in an organizational setting could be assisted with improved information flow. However, demonstrating the validity of the joint production principle to organizational development to assist with rights-based actions requires further empirical work.

First, employing the joint production principle centralizes information (its generation, dissemination and use) in making rights-based claims. As demonstrated above, information asymmetries need to be overcome in order to strengthen people's claims in a sustainable way. Current rights-based theories of change that draw on the UN framework are incomplete in their guidance on the incorporation of information. By building in the principle of joint production, information utilization can be mainstreamed into programme design. Joint production fills a gap in the current theories of change of the UN framework.

Second, applying three tests at the heart of the principle of joint production aligns the analytical scope of the programme by focusing on issues of most salience to the interest groups involved in decisions. This means that analysis focuses on defining action around the problems of the largest size that happen regularly, as opposed to conducting far-reaching analysis into all potential claims. Focusing on the definition of salient questions with a range of interest groups could be criticized, as it could potentially marginalize the voice of the poor. This is a risk, which needs to be weighed against the necessity of persuading the powerful to buy in to a change process. This risk is also partially mitigated by the improved spread and use of information. Enabling poor people to apply information to a problem can increase their power, as demonstrated by ActionAid's Reflect methodology (Duffy *et al.*, 2009). In South Africa, constitutional guarantees secure a framework to assist rights-based discourse. In countries where democratic institutions are under attack, or with limited provision for the voice of marginalized people, the risk is higher. More work is required in this area.

Third, focusing on salient issues identified by multiple interest groups helps to prioritize the most relevant human rights concepts. Vernacularization of human rights language (Merry, 2006; Sarfaty, 2010) can then be applied in a focused manner that strengthens people's ability to interpret the claim in relation to a limited number of issues. In this regard, useful examples can be drawn from the TAC's work in operationalizing the right to health (Heywood, 2009).

Joint production is not an easy option. Expertise is still required in a number of fields, such as community development, information synthesis and vernacularization of human rights. However, through applying the joint production principle, development of a theory of change becomes less amorphous and easier to perceive because of the prioritization of actions and information flows with important interests. This can help to bridge accountability gaps by preparing the ground for shared conversations on action. In effect, joint production can support partnerships and the definitions of multi-dimensional human rights strategies (see chapter 1 in this volume) by helping to prioritize, sequence and divide labour between various actors, with information forming the ground for shared bases for interaction. In the ongoing redesign of Impilo, this means:

- confirming what information is needed for people to better access health services
- more focused conversations with interest groups as to salient issues with which Impilo needs to engage
- development of new partnerships with organizations that are already considered legitimate in the monitoring of government services
- engaging new skills in community mobilization and capacity development.

Conclusion

The experience of implementing Impilo indicates some areas where rights-based theories of change can be improved. The argument here is that the joint production principle tempers some of the concepts that are intangible and helps to define actions between rights-holders and duty-bearers through improved information flows. Given the earlier descriptions of accountability deficits in South Africa, this is a promising area to improve the interface around claims. More work is required to understand better how joint production links to improving various forms of capacity, and to see how it can work where people's voices are systematically closed out.

Impilo, it should be conceded, was a project principally concerned with information. The conclusions arrived at here could therefore be contested, as information would always appear important. However, that does not mean that current rights-based theories of change have considered the role of information adequately. A pool of information is required to make rights-based claims and interface around the definition of quality service delivery. Information helps to lubricate discourse and enables people to make their own claims based on what they value. Without expanded information, people might not know what services there are, let alone be able to evaluate their performance. Joint production provides a useful principle that places the use of information at the heart of the rights-based theories of change.

This does not mean that we abandon all rights-based concepts. Rather, we should change roads. Another route to a useful rights-based analysis is possible by prioritizing tangible and categorical mechanisms that relate to the strengthening of people's interactions with government – joint production being one proposed principle to guide action. Initiating processes in which people and the health system work together to understand better each other's needs and limitations helps to define directed interventions. Asking if people are respected by the health system or if the health system respects a person has all the solidity of air.

Note

1 This chapter is the result of a number of years' thinking about and working towards a rights-based programme, as a result it is the

product of many hands and heads. I am particularly grateful to Patricia Rogers, Ray Rist and Barry Munsolow for contributing towards the thinking on this chapter and commenting on earlier versions. Funding for completion of this research came from the University of the Witwatersrand, where I currently work. Thanks also to the International Development Evaluation Association (IDEAS), which gave me a forum to discuss the ideas. The work on developing the Impilo programme theory drew as well from many heads, foremost Debbie Heustice and Catherine Jenkin. In addition I would like to thank Ravi Ram, Rita Sonko, Addis Berhanu, Melusi Ndhlalambi and Phathisiwe Ngwenya for support on this project. In the field, Impilo benefited from all the staff of the AMREF Bana Barona programme, who helped develop and implement the system, especially Nonhlanhla Mthimkhulu, but with support from Penina Ochola, Juju Mlungwana, Ntombi Mabindisa, Lovemore Mhuriyengwe and Itumeleng Masia. Thanks also go to all the co-contributors to this book for their comments during various meetings. Finally, thanks to all staff from the community organizations, who are too numerous to mention, but with special recognition to coordinators and data capturers who have stuck with the project in Ndumo Community Care Centre and Ubomo Home-based Care: Thembelihle Qwabe, Dudu Mhlanga, Joyce Mdluli and Phumzile Vilakazi. Any mistakes in the chapter remain my own and do not necessarily represent the views of all of those involved in the development of the system.

References

Akerlof, G.A. (1970) 'The market for "lemons": quality uncertainty and the market mechanism', *Quarterly Journal of Economics*, 84 (3): 488–500.

Benington, J. and Moore, M.H. (2011) 'Public value in complex and changing times', in J. Benington and M.H. Moore (eds), *Public Value: Theory and Practice* (London: Palgrave Macmillan), pp. 1–30.

Bloom, G., Standing, H. and Lloyd, R. (2008) 'Markets, information asymmetry and health care: towards new social contracts', *Social Science & Medicine*, 66 (10): 2076–87.

CESCR (2000) 'General Comment No. 14: The right to the highest attainable standard of health' (Geneva: Social Committee on Economic, Social and Cultural Rights).

Chen, H. (2005) *Practical Program Evaluation: Assessing and Improving Planning, Implementation and Effectiveness* (Newbury Park, CA: Sage).

Clark, W.C., Mitchell, R.B. and Cash, D.W. (2006) 'Evaluating the influence of global environmental assessments', in R.B. Mitchell *et al.* (eds), *Global Environmental Assessments: Information and Influence* (Cambridge: Massachusetts Institute of Technology), p. 1.

Coats, D. and Passmore, E. (2008) *Public Value: The Next Steps in Public Service Reform* (London: Work Foundation).

Day, C., Monticelli, F., Barron, P., Haynes, R., Smith, J. and Sello, E. (2010) *District Health Barometer Year 2008/09* (Durban: Health Systems Trust).

Donnelly, J. (2003) *Universal Human Rights in Theory and Practice*, 2nd edn (Ithaca, NY/London: Cornell University Press).

Duffy, M., Fransman, J. and Pearce, E. (2009) 'Review of 16 Reflect evaluations' (Johannesburg: ActionAid), www.reflect-action.org/node/68

Funnell, S.C. and Rogers, P.J. (2011) *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models* (San Francisco: Jossey-Bass).

Heywood, M. (2009) 'South Africa's treatment action campaign: combining law and social mobilization to realize the right to health', *Journal of Human Rights Practice*, 1 (1): 14–36.

Jonsson, U. (2005) 'A human rights-based approach to programming,' in P. Gready and J. Ensor (eds), *Reinventing Development? Translating Rights-Based Approaches from Theory into Practice* (London: Zed Books), pp. 47–62.

KwaZulu Natal Department of Health (2010) *Annual Performance Plan 2010/11–2012–2013* (Durban: KwaZulu-Natal Department of Health).

Lee, K.N. (1999) 'Appraising adaptive management', *Conservation Ecology*, 3 (2): 3.

Madsen, H.L. (2007) 'Exploring a human rights-based approach to the evaluation of democracy support', in P. Burnell (ed.), *Evaluating Democracy Support: Methods and Experiences* (Stockholm: International Institute for Democracy and Electoral Assistance).

Merry, S.E. (2006) 'Transnational human rights and local activism: mapping the middle', *American Anthropologist*, 108 (1): 38–51.

Nyamu-Musembi, C. and Cornwall, A. (2004) *What is the 'Rights-Based Approach' About? Perspectives from International Development Agencies*, Working Paper 234 (Brighton: Institute of Development Studies).

OHCHR (2006) 'Frequently asked questions on a human rights-based approach to development cooperation' (Geneva: Office of the High Commissioner for Human Rights). www.ohchr.org/Documents/Publications/FAQen.pdf

Packard Foundation (2010) *Linking Knowledge with Action* (Los Altos: David and Lucile Packard Foundation). www.packard.org/wp-content/uploads/2011/01/ScienceStrategy-Linking-Knowledge-with-Action1.pdf

Sarfaty, G.A. (2010) 'Internal conflict over the World Bank's empirical approach to human rights', in K.M. Clarke and M. Goodale (eds), *Mirrors of Justice: Law and Power in the post-Cold War Era* (Cambridge: Cambridge University Press).

Sen, A.K. (2009) *The Idea of Justice* (Cambridge, MA: Harvard University Press).

Stiglitz, J.E. (2002) 'Information and the change in the paradigm in economics', *American Economic Review*, 92 (3): 460–501.

SWOP (2011) *The Smoke that Calls: Insurgent Citizenship, Collective Violence and the Struggle for a Place in the New South Africa* (Johannesburg: Society Work and Development Institute, University of Witwatersrand).

UNDP (2003) *UN Statement of Common Understanding on Human Rights-Based Approach to Development Cooperation and Programming* (New York: United Nations Development Programme).

World Bank (2010) *Accountability in Public Services in South Africa* (Washington, DC: World Bank).

Appendix E: Porter, S. (2017). 'A Framework for Identifying Entry Points for International Development Evaluation to Enable Responsive Government Policy'. In D. Podems (Ed.), *Democratic Evaluation and Democracy: Exploring the Reality*. Charlotte, NC: Information Age Publishing.

CHAPTER 2

A FRAMEWORK FOR IDENTIFYING ENTRY POINTS FOR INTERNATIONAL DEVELOPMENT EVALUATION TO ENABLE RESPONSIVE GOVERNMENT POLICY

Stephen Porter

ABSTRACT

International development evaluation, which prioritizes values linked to poverty eradication, can help policy respond to poor people and marginalized people's concerns. Power imbalances, however, mean that although evaluation that applies the standards of international development has mechanisms to contribute to the improved responsiveness of government, the interests of poor people are often excluded from the process (House & Howe, 1999). In this chapter, a framework that utilizes principal-agent concepts is applied, drawing on five case studies, and some potential entry-points for evaluation to support policy processes to respond to poor and marginalized people are

Democratic Evaluation and Democracy: Exploring the Reality, pp. 17–36

highlighted. The discussion provides initial evidence of the applicability of the principal–agent concepts to the implementation of international development evaluation practice so that it can better enable responsive policy decisions.

Governments are responsive when they "make and implement policies that citizens want" (Powell, 2004, p. 62). Government that is responsive to citizens, who are considered its political equals, is the key characteristic of democracy (Dahl, 1973). When a government changes policy or implementation as a result of taking into account the interests of the whole of society, rather than its own narrow interests, through procedural mechanisms it demonstrates democratic responsiveness (Fukuyama, 2014). Procedural accountability mechanisms, such as free and fair elections, corruption monitoring bodies, and the rule of law work collectively to produce responsiveness (Diamond & Morlino, 2005). Forms of state that are not democratic, such as monarchies and those that have one party rule, can also be responsive through more limited forms of accountability than those found in democracies. Improving the functioning of accountability mechanisms through applying the standards of international development evaluation, which prioritizes values linked to poverty eradication, can support a government's responsiveness to citizens in whatever form of state by generating channels to enable consideration of otherwise excluded perspectives.

It is argued in this chapter that international development evaluation offers a substantive tool to enable responsiveness. The practice of evaluation,[1] broadly, helps to define the merit or worth of an intervention according to a set of defined values, principles and standards. International development evaluation, specifically, is required as part of the implementation of the Sustainable Development Goals, whose mandate and reach is a globally agreed. The values of these Goals and consequently the standards of international development evaluation, include inclusivity, eradicating poverty and strengthening universal peace in larger freedom (General Assembly, 2015). Prior to the Sustainable Development Goals, a range of groups have elucidated guidelines for international development evaluation practice. These include, but are not limited to: The Organization for Economic Cooperation and Development—Development Assistance Committee's (OECD-DAC), which produced the *Criteria for Evaluating Development Assistance* and *Quality Standards for Development Evaluation*; and the International Development Evaluation Association's *Evaluator Competencies*.

Power imbalances, however, are an impediment to the contribution that international development evaluation can make to accountability processes. Power imbalances mean that although evaluation has the possibility

to be an important contributor to poverty reduction efforts, the interests of the poor and marginalized are often excluded from the process of evaluation (House & Howe, 1999). Practical mitigations for the distorting effect of power have been defined for those who *conduct* evaluations, such as, opening space for dialogue with the poor and being inclusive of their voice in evaluation processes (House & Howe, 1999; Patton, 2008). Less guidance is available, however, to support those that *manage* development evaluations and seek to embed them institutionally.

This chapter seeks to make a contribution to evaluation practice by defining a framework that applies principal-agent concepts to support the identification of entry points in accountability mechanisms to improve responsiveness of government to the values and ends of the Sustainable Development Goals, which prioritize poverty eradication. The framework is based upon a knowledge base developed in undertaking five exploratory case studies (Porter & Feinstein, 2014).

DEVELOPMENT EVALUATION, POWER AND A PRINCIPAL-AGENT FRAMEWORK

Shadish, Cook, and Leviton (1991, p. 46), in their classic text *Foundations of Program Evaluation*, argue that "no prescriptive [value] theory is widely accepted … all prescriptive ethics are under justified, and selecting one involves trade-offs about which few stakeholders agree." In this assessment they are broadly correct. The world is a place of diverse values where knowledge is socially constructed. This means that not all evaluation can promote democracy, as it is not something universally valued. Though democracy is not widely accepted, the bases of international development evaluation are, curiously, widely accepted; the UN General Assembly through the Sustainable Development Goals affirmed the values and standards of international development evaluation. In the real world, articulated intent and actual implementation diverge, however. Those with power often intentionally exclude important issues in policy processes, such as, inclusion, human rights and environmental sustainability. International development evaluation has a mandate and institutional mechanisms to highlight these exclusions.

Translating the mandate of the Sustainable Development Goals that prioritizes a range of values linked to poverty eradication is challenging, however. Those who are in power circumscribe how evaluation processes work; their interests define key questions, data collection methods and the space of allowable judgements. The poor and marginalized are often excluded from these processes because accountability mechanisms, and by extension development evaluation processes, are not sufficiently interested

or connected to their perspectives. If the diverse interests of poor people are excluded from processes, development evaluation cannot support policy responses in accordance with their values.

There are three dimensions of power (Lukes, 1974, pp. 16–29) with which international development evaluation can interact in policy processes. The first dimension is the power to make people do something: "*A* exercises power over *B* when *A* affects *B* in a manner contrary to *B's* interests." The second dimension is the power to shape the political agenda to organise an issue into or out of the political domain. The third dimension is the power of socially structured and patterned behaviors of groups; a group may not even be conscious that their interests are contrary to the prevailing agenda.

The commissioning of an evaluation by powerful groups who intend to use the findings in a decision is an outright act of power. Development evaluation can also be a process that produces alternatives for a policy agenda. An evaluation can be commissioned in order to help raise an issue. The definition of alternatives "*is the supreme instrument of power*; … [they] who determine what politics is run the country, because the definition of alternatives is the choice of conflicts, and the choice of conflicts allocates power" (Schattschneider, 1960, p. 68). A development evaluation process is an act of power, especially when it prioritizes understanding the interests of those with limited power, because it identifies an issue, measures it, interacts with the interested, passes judgement and, on that basis, proposes alternatives. An evaluation undertaken in accordance with international development standards may even be able to help confront false consciousness, but this dimension of power is not considered further in this chapter.

Governments can articulate intent to achieve the values of the Sustainable Development Goals, such as, inclusion and human rights, but undertake actions that are repressive. In this case, the definition of alternative actions can be a significant challenge to those with established power, as it contradicts their actions. Commissioning an evaluation in a way that offends too many interests will mean that the implementation of change is likely to be more challenging. Consequently, powerful interests often need to be navigated rather than just challenged, through given entry points.

Accountability mechanisms are important entry points to challenge powerful interests. Accountability mechanisms at their core are a way to constrain the misuse of power (Lindberg, 2013). In states, whether democratic or not, accountability mechanisms exist that are bureaucratic, informal, or patron-client orientated. Accountability mechanisms provide possibilities for development evaluation to engage with powerful interests that set agendas, provide sanctions and produce an official acceptance of a position. If evaluations are to be commissioned that can support interventions to respond to people, especially the poor and marginalized—and give

prominence to their perspective—then a framework is required to identify amenable accountability mechanisms.

A useful frame for investigating accountability mechanisms are principal-agent concepts as they are "at the heart of public sector accountability research" (Schillemans & Busuioc, 2015, p. 191). Applying principal-agent concepts focused on accountability mechanisms helps to understand the institutional dimension of evaluation in development, helping to identify balances that need to be struck between different entry-points (Picciotto & Wiesner, 1998). Principal-agent relationships have been developed and applied as a conceptual framework by political scientists (Lindberg, 2013) and economists (Stiglitz, 2002) to help unpack power relationships. Given this academic background, their application offers a developed body of literature that can be integrated into evaluation approaches.

Working through principal-agent relationships helps to analyze where power is located, contracted and transferred. This section now outlines a framework to apply principal-agent concepts to international development evaluation practice by defining key terms and relating them to existing literature in evaluation. The section concludes by providing examples to show how principal-agent concepts can be applied to thinking about implementing development evaluation.

Principals

In a liberal democracy a principal (citizen) transfers power to an agent (politicians) to act on his/her behalf. In turn politicians become principals by transferring some power to civil servants (agents), who then become principals in the delivery of duties to citizens. The transfer of power between principals and agents is held-in-check by formal or informal accountability mechanisms that can evaluate performance and potentially apply sanctions (Lindberg, 2013). In non-democratic contexts, principal-agent relations are often mediated through processes that do not involve direct representation and are unlikely to prioritize transparency.

Principals are the demand side of evaluation and in this paper broadly define the prevalent configuration of the authorizing environment of the state. The authorizing environment in the context of this paper consists of people in formal decision-making power within accountability mechanisms.[2] When decision makers require or are influenced to ask for evidence, then an actual, latent or potential demand arises for evaluation. The demand is latent if the decision maker is not aware that evaluation can be a source of evidence. The demand is potential if there is an awareness of evaluation, but resources to fund the evaluation are lacking.

Principals have different configurations of demand for evaluation as well as interests. For example, if the executive focuses on improving performance through monitoring, then there is a latent demand for evaluation. Meanwhile, the legislature might be focused on getting the executive to account for the effective use of funds. In this case a potential demand for evaluation is generated if they do not have the resources to fund evaluations.[3] The tensions that exist between different role-players in the authorizing environment generate possibilities to elicit actual demand for evaluation. Cabinet, Parliament, political parties, civil society, senior civil servants, and the media are some of the main principals who demand evaluation.

When principals, who play a leading role in the state, articulate demand for evaluation, the findings of the evaluations can play a potentially important role in policy processes as these principals have actual power over processes. Informational asymmetry exists between principals and agents who implement policy, however. Principals may have positional power, but they have limited levers to ensure action. Evaluation can be important as it provides information on implementation and can bypass layers of interests and present to principals a reality of implementation based on different values.

Evaluation that is channelled through a locally owned accountability mechanism improves the legitimacy of findings and the potential for action. Principals are more likely to use findings from an evaluation when demand for evaluation processes arises from within a country's politics. The importance of demand arising internally to a system has been recognized within the literature on evaluation (Bemelmans-Videc, Rist, & Vedung, 2003; Boyle & Lemarie, 1999; Chelimsky, 2006; Mackay, 2007; Picciotto, 1995; Wiesner, 2011). When demand for evaluation is internal (endogenous) to a country system, but the principal that takes on the role of championing the evaluation process is outside the normal policy process and interacts with potential sources of disagreement, power is exercised: a potential conflict has been chosen, an agenda set.

In international development evaluation, donors often play a role and can choose to exercise demand for evaluation outside of the accountability structures (exogenously) of the partner country, accounting to the authorizing environment in their home countries. Donors can, if they choose to, interact with a partner country's political power structures when commissioning evaluations or when attempting to support their findings.[4]

Those seeking to advance international development evaluation processes to respond to the poor will need to identify the intersection of stakeholders that are willing to seek change. They will need to operate in a politically savvy manner and identify actual, latent, and potential demand for change within the larger system of power. They need to be able to

identify actors who are interested and want to frame and resolve a problem and value the perspective of the poor.

Agents

The supply-side of evaluation processes consists of agents who undertake evaluations. Evaluation agents (evaluators) ought to have more knowledge than other role-players in technical areas of evaluation. For example, they should have in-depth knowledge of evaluation methodology. If incentives and regulation are weak, a lower standard of service might become prevalent. This is because evaluators benefit from information asymmetries, they know more about the work they conduct than those managing the evaluation. Evaluation agents may work together to overcome inadequate incentives by creating shared regulation mechanisms to improve practice.

Evaluation agents need to be competent in undertaking the collection, capturing and verifying data, and interacting with the system for managing, designing and disseminating credible evaluation. Technical capacity on the supply-side, however, is insufficient to ensure the use of evaluations. Credible data produced by technically sound agents does not mean that these data are relevant to the existing political context. The strength of evaluation supply is predicated not only on its technical know-how, but also on its ability to relate to existing institutions, particularly accountability mechanisms. In other words, evaluations should not just be credible, but also answer important questions and be legitimate in the political context. Evaluators who are seen as legitimate are more likely to get access to important stakeholders and will know how to frame recommendations so that they are accepted within the context. Universities, think tanks, consultants generally conduct evaluations as supply-side agents. Each of these can play different roles when thinking through possibilities within a given context.

Intermediary Managers

Evaluation intermediaries are entrusted to act on behalf of principals that they are supposed to serve in their work between principals and agents. They manage evaluation implementation, and liaise with commissioners of the evaluation process.

An illustration of intermediaries is the Senior Responsible Owner in U.K.'s Department for International Development (DFID) who oversees the implementation of projects. Senior Responsible Owners manage or delegate the management of evaluations, for example, to evaluation advisors. The evaluation advisor and the Senior Responsible Owner are

intermediaries. For both, there is someone higher in the authorizing environment that accounts for the performance of a portfolio of projects, but is not responsible for day-to-day project implementation, does not manage the evaluation, and puts in place mechanisms that generate demand for evaluation. Often, intermediaries that manage international development evaluations need to weigh input from different principals both inside and outside their government. Intermediaries in these positions are able to develop knowledge of institutions and policy that allows them to influence demand and evaluation implementation, potentially, in accordance with their own agenda.

Good intermediaries need to understand the technical conduct of evaluations, but also need distinctive skills in helping to relate evaluations to accountability mechanisms, such as select committees, budget allocation processes, and sector management committees. Those managing evaluations within a country system are often connected to demand through accountability mechanisms, which can be public good orientated or patronage based. Intermediaries who are external to an authorizing environment, such as evaluation advisors in DFID, need to find a channel to connect to demand through in-country structures. With such a channel, they may be able to improve the use of evaluation within the partner country, as opposed to applying findings only to the donor's program.

In summary, the model presented here deviates from those presented elsewhere in the evaluation literature by highlighting intermediaries. The importance of intermediaries in government change processes has been recognized in literature outside of evaluation (see, for example, Klerkx & Leeuwis, 2008). There are a number of other potential actors who could act as intermediaries and help evaluation practice, for example, associations of professional evaluators and grant funded innovators.[5]

Example Application to Development Evaluation

The conceptual framework of principals, agents, and intermediaries could be expanded. The scheme above enables insights on the role of different stakeholders in evaluation processes and their relative positional power. Understanding the configuration of power, who is delegating and transferring power to whom, helps to unpack the "social relations, particularly the power relations, that mutually constitute the production, distribution, and consumption of … [policy]" (Mosco, 1996, p. 24). A principal-agent lens enables those advancing evaluation systems, often intermediaries, to pay attention to the forces and processes at work and how they affect evaluation.

Two examples below illustrate how the principal–agent distinction can be applied to evaluation problems.

The principal-agent distinction helps to understand how power is configured within the authorizing environment. While principals transfer power to an agent, there are issues with who monitors the transfer of power. In countries where development aid is delivered, power is often centralised in a strong executive or elite group. Knowing how the elite group operates helps to identify where accountability mechanisms work for private, rather than public interests. A legislature, for example, may be systematically weak because it does not offer a genuine constraint on executive power. In this context, the bodies that report to it, perhaps a public service commission, would also have limitations. This means that the search for an appropriate body with the power to commission evaluations would have to be sought in either the central executive or other parts of the state, such as, civil society where there is capacity to challenge powerful interests. Examples of power configurations in relation to international development evaluation systems are presented in the next section of this paper.

The principal-agent distinction can also help evaluation managers and other intermediaries think about how they relate to those conducting evaluations, especially in terms of market failure. In many developing countries an array of demands are emerging for evaluation (Porter & Goldman, 2013). In an emerging market where principals and intermediaries lack experience there is the potential for informational asymmetry, which can lead to poor service value. George Akerloff's (1970) article *The Market for Lemons* highlights the importance of informational asymmetry in the quality of the delivery of goods and services. The owner of a car (agent) has more information than the potential buyer (the principal), as a result incentives are produced to up the price of the car and conceal defects. In the same way, where an evaluator has more information on the kinds of evaluation that are possible and can exaggerate his/her own experience or conceal other approaches, then incentives are produced to offer lower quality services. Because of a lack of knowledge the principal and intermediary do not necessarily have the background to adequately demand and manage an evaluation, meaning that lower quality evaluations become passable. In this context an intermediary in the form of an evaluation manager can draw in additional support by implementing expanded governance structures and quality assurance mechanisms to improve information symmetry.

In summary, the principal-agent distinction offers a useful way to conceptually delineate the power relations among different stakeholders involved in evaluation. Different groups, whether principals, agents or intermediaries, interact with politics and power in different ways. The roles they play in policy processes can change according to context. Working

through how power is practiced helps to identify potential issues and opportunities for development evaluation processes.

EMERGING PATTERNS OF PRINCIPAL-AGENT RELATIONS FROM CASE STUDIES

Applying the principal-agent conceptual framework within five case studies helps to identify entry points for international development evaluation in different power configurations (Porter & Feinstein, 2014). For evaluation to respond to the poor, a stakeholder is required who is willing to exercise power in the interests of the poor. An accountability mechanism is needed that represents genuine concern about the poor. And agents are needed who are willing to use their power or choose potential conflicts with interests that are adverse to some of the preferences of the poor. Those with power in an authorizing environment are not often poor people, but the authorizers can be interested in improving the situation whether for altruistic or selfish reasons.

A manager of a development evaluation or another party interested in advancing evaluation practice is likely to have a partial view of the local anatomy of power and a very limited ability to affect policy choices. Even with a partial view and limited influence, the manager who identifies a champion and supportive accountability mechanisms can implement an evaluation process that is influential in decision making. The configurations and illustrations discussed below demonstrate emerging patterns in Ghana, Ethiopia, Malawi, Rwanda, and Zambia and the application of a principal–agent conceptual framework in supporting the development of evaluation systems.

Each of the case countries, at the time of writing of the cases in early 2014, had enjoyed a sustained period of economic growth and a good degree of fiscal and monetary stability. Ghana has a variety of institutions that are developing a democratic outlook, although power is still exercised by a narrow elite. Malawi and Zambia have more limited democratic procedures. In both, there are also limitations in the power of the government to regulate powerful vested interests. In Ethiopia and Rwanda, powerful central elites are organized in party structures that emerged through violent conflict. It these countries there is a good degree of ability to implement policy, but democratic spaces are limited.

The study was carried out through a combination of desk review, including an analysis of existing evaluation/evaluative research products, and direct semi-structured interviews with a selection of informants across critical stakeholder groupings. In total, people in 77 agencies were interviewed as part of this study. The primary data collection for the case

studies was carried out from April to June, 2013. Further verification and analysis took place between June and December, 2013.

In applying the principal-agent conceptual framework, the study was undertaken through the following overlapping stages: Established study commitment and support from key stakeholders; collated and analyzed primary and secondary data and information about the evaluation system; conducted a series of interviews with actors that fall within the categories of principal, intermediary and agent; reviewed additional secondary information; and finally, produced a country case study paper.

Identified Configurations and Their Constraints[6]

There are two overarching political configurations that serve as useful reference points for understanding power relations and opportunities in the five case countries: (1) *neopatrimonial*, in which public affairs are to a large extent subservient to entangled personal and private interests; and (2) *developmental patrimonial*, in which a central ruling elite command and promote long-term development in the interests of maintaining power (Booth et al., 2005; Booth & Golooba-Mutebi, 2012). Elements of (3) *liberal democracy* can be identified within all the case countries. In a liberal democratic structure, transparent rule-based processes relate to accounting mechanisms, where the diverse values of poor people could be considered amongst a variety of other interests. Even though structures that are transparent, rule-based and inclusive are rare in the case countries, mechanisms were identified where they could be further refined or developed. All three logics offer distinct entry points, operate to different degrees within a state and suggest distinct approaches for those undertaking international development evaluations. Even so, in each case country one logic was normally prevalent.

In the *neopatrimonial* authorizing environments of Zambia, Malawi, and to a lesser extent Ghana, the informal forces that shape decision-making or implementation processes are opaque and hard to reform through direct intervention. An authorizing environment with neopatrimonial characteristics:

> On the one hand [has] … a formal administrative structure governed by rules and underpinned by law. On the other hand, much of the actual operation of public affairs is dictated by a different set of principles. State resources, bureaucratic positions, and the power to allocate rents, provide services, and determine policies and their beneficiaries are captured by personal or private networks in the hands of dominant patrons. Thus, instead of being governed by explicit objectives and legal rules, it is effectively an apparatus

> serving the interests of the particular groups and individuals that have captured it. (Booth et al., 2005)

A neopatrimonial logic can still mean that development in terms of economic growth is achieved, but in a system where the main benefits accrue to the powerful.

A neopatrimonial logic is prevalent in the authorizing environments of Zambia and Malawi. The main objective of a range of principals is the consolidation of power within existing private networks. In this logic agents order or disorder a policy structure or process so as to enable personal, group, or institutional benefit to be derived. For example, disorder of information management, such as, poor record keeping, can represent a strategy to undermine accountability mechanisms, rather than a lack of capacity. Similarly, order, such as a public service reform programme can provide an opportunity for consolidating power, diverting and accumulating resources. On a large enough scale patron-client accountability mechanisms predominate the primary functions of the state rather than the public good. The consequence of this is that policy becomes difficult to *actually* change as interests mediate implementation and powerful interests assert control over reform to better achieve their objectives. Ghana, meanwhile, had a range of differing authorizing environments varying between liberal democratic and neopatrimonial, although neopatrimonial logic appeared to be prevalent.

For international development evaluation this means that apparent entry points could be a façade for the accumulation of power. The existence of competing interests, however, does allow for alternative entry points for evaluation as well as for the promotion of competing users of evaluation evidence. The challenge is to identify which of these is most likely to be interested in being responsive to poor people. The consequences of the different neopatrimonial logics in identifying entry-points for evaluation are outlined in the next section.

In *developmental patrimonial* authorizing environments policy-making is centralised and strict limits on policy influence are imposed. Developmental Patrimonialism can be recognized when the:

> Ruling elite acquires an interest in, and a capability for, managing economic rents in a centralised way with a view to enhancing their own and others' incomes in the long run rather than maximizing them in the short run. (Booth & Golooba-Mutebi, 2012, p. 381)

In this logic the political establishment seeks to maintain power and build support through demonstrating an ability to provide development, while often subordinating processes that give rise to political competition in the way accountability mechanisms function. Developmental patrimonialism

often curtails dispersed corruption as a means of legitimating and consolidating the position of elites linked to ruling political parties; by extension open debate of policy options is also repressed. Patronage in this structure is centralised, often through companies that relate to the dominant party (Booth & Golooba-Mutebi, 2012). This option for maintaining power is selected over outright repression. In such environments criticism of policy is not just about contesting a decision-process, but a challenge to the development vision of the elite. This logic reduces the role of international donors in defining policy. If the ruling elite sees a defined policy change as helping to achieve development in a manner that reinforces their control they will pursue that objective. A policy change is unlikely if it directly undermines the long-term control of the elite. The case studies identified this kind of decision making as prevalent in Ethiopia and Rwanda.

Developmental patrimonialism offers opportunities for development evaluation when a powerful section of the elite has an interest in improving the effectiveness of the policies they have chosen. The challenge of this is that, while they are open to change the implementation of programmes, the definition of policy remains in their domain. This means that their responsiveness to the poor is bounded; evaluating the value of a policy is more difficult than providing feedback from the poor on the implementation of the policy. Any challenge to the centrally managed patronage structure is likely to remain off-limits.

POTENTIAL ENTRY POINTS IDENTIFIED THROUGH APPLICATION OF A PRINCIPAL-AGENT LENS

This section presents some of the findings of the case studies and illustrates the potential application of the principal-agent lens in identifying entry points in the authorizing environment for evaluation practice. In all countries there are some potential entry points amongst both principals and agents. The purpose of the analysis in this section is to highlight issues that arise within the different stakeholder groupings of principals, intermediaries and agents and illustrate the application of principal–agent concepts.

Principals

In the case studies there are marked similarities between principals in developmental patrimonial and neopatrimonial environments. The executive is the dominant branch of government in all of the case countries. For evaluation to be taken seriously an entry point is required within the executive, whether in central or line ministries.

Across the studies there is actual demand for monitoring and on occasions, in developmental patrimonial environments, evaluation focused on the merits of implementation, but not questioning the worth. Research is less threatening than evaluation; it provides evidence about a theory. Evaluation meanwhile strives to present a reality from a value perspective, which may challenge centrally held beliefs that could raise unnecessary challenge to elites.

In developmental patrimonial environments, although monitoring remains dominant, there is some openness to research being used in policy processes. In Rwanda, for example, informants from government and civil society report that there is an active demand for using evidence and monitoring information to inform the executive. If evidence is presented that is critical of a policy decision, especially on sensitive issues, the discussion might be rebuffed. This was reported to have happened in Rwanda when an evaluation of the effects of universal access to nine to 12 years of basic education on school performance started to show negative results. The general openness to evidence and monitoring information within the executive of Rwanda is not the same as an active demand for international development evaluation, which directly incorporates a variety of value positions on an intervention. A savvy development professional could introduce some deeper evaluation practices that promote responsiveness around the implementation of policy.

In the neopatrimonial environments, monitoring is also the dominant form of performance assessment utilized by the executive. In Zambia, at the level of the executive there was a demand for regular monitoring updates, rather than evaluation. In Malawi demand has been weak and inconsistent, in the case study this is argued to be as a result of a deeply rooted historical culture of patronage and fluid policies.

In contrast, within Ghana the executive arm of government has a quite strong latent demand for evaluation in a liberal democratic milieu. Over a number of years the Presidency of Ghana has tried to establish units to monitor and support implementation. The Policy Evaluation and Oversight Unit has been replaced with a reconfigured Delivery Unit. The demand of the current president for improved delivery is acting as a latent demand for evaluation, which some advisors are seeking to catalyze into actual demand. Such units linked to the executive can provide strong entry points for evaluations and in some conditions may have an interest in centralizing the perspective of the poor in policy processes.

Across all national legislatures in the case studies the appetite for and even the capacity to recognize the value of independent research and evaluation for oversight is low. Evidence emerges from the cases, however, that indicate that developmental patrimonial environments undertake more serious oversight processes through audit reports than legislatures

that function in a neopatrimonial process. Within Rwanda there is emerging demand for more information on government activities through audit reports especially on corruption issues. For example, in Rwanda parliament's longest serving member at the time of this study heads the Parliamentary Accounts Committee, which was created in April 2011. In neopatrimonial environments the limited reach of Parliament's oversight is reflected in their inability to follow-up on the auditor general reports. In Ghana and Zambia there is specific evidence on the inability to curtail malfeasance where the reports of the auditor general are not followed-up on. In Malawi the indications are that the Public Accounts Committee of the Assembly engages periodically with information presented by the audit process. The potential of the legislature as intelligent users of evaluation in the case countries was vast as they can conduct oversight of all areas of government, but ultimately mediated by the power of the executive and the overall political logics of the state.

Demand for evaluation from civil society and meaningful use of evaluation results are complicated. In neopatrimonial environments there are some channels for civil society to utilise the media for debate. Across all countries entry points in civil society could be identified for international development evaluation generally amongst older civil society actors that have developed their legitimacy in the political context over time.

The five studies show that development partners (donors) dominate the actual demand and implementation of evaluation studies. In interviews, evaluators conceded that most evaluations were commissioned and managed by development partners. Evaluation practice has emerged in those sectors where donors focused their evaluations, such as health and education. It should be noted that some evaluations commissioned by development partners do align with country-led questions and that donors do make resources available for evaluative activities, for example, DFID's funding of the President's Delivery Unit in Ghana.

Some evaluations undertaken with development partner resources do align to country-led practices. The studies show that it is too simplistic to consider development partner demand for evaluation as narrowly focused upon their own interests. On occasions, when government takes an interest, development partner evaluations can lead to shifts in implementation. For example, in Malawi the government used recommendations from an evaluation of the National Cash Transfer Programme that was commissioned and led by development partners. In Rwanda, an influential randomized control trial has been completed in relation to performance financing in health (Basinga et al., 2011). In both these instances there appears to have been an alignment of latent demands for information on a policy issue of interest within the political configuration of the country. This example illustrates that when evaluation aligns to some of the existing interests it

can be used in a manner that changes policy and responds to the interests of the poor and marginalized.

Intermediaries

Central Agencies (Ministries of Finance and Planning Commissions) are custodians of policy intent and support the allocation of resources. In reality, the power of the central agencies is mediated by political power. Developmental patrimonial environments tend towards stronger central agencies that can reallocate resources in alignment with the desires of the central elite. Meanwhile, in neopatrimonial environments in central agencies, resource allocation is the result of informal policy processes, which means that central intermediaries are institutionally weaker and less able to align resources to strategic intent. Line Ministries, such as Health, may in neopatrimonial contexts be more open to evaluation activities that seek to further understand the effectiveness of interventions from the perspective of poor people.

In all the case countries, the organizational arrangements for monitoring and evaluation were articulated in the National Development Plans. These usually place a Central Agency as a coordinating body. These arrangements have historically in the main focused upon monitoring, especially the development of hierarchical monitoring systems that feed into annual progress reports, for example, the national development plans of Ghana and Zambia. These plans do provide entry points for evaluation in spite of the focus on monitoring that could legitimately highlight the perspective of the poor in development plans. In Zambia, for example, ranges of evaluations were undertaken for the fifth national development plan. In Rwanda, evaluations have been commissioned through the Ministry of Finance though often working with development partners.

In order to respond to and formalise potential and latent demands into actual demands for evaluation that prioritizes the poor, intermediaries need to take opportunities to improve the evaluation management function. In developmental patrimonial environments this could be accomplished within the Ministries of Finance given their strong technocratic role. Within neopatrimonial environments, which have weaker central ministries, the approach might identify assets in a central ministry, such as Finance, but also work across the system, such as Ministries of Health.

Agents

There are few apparent differences between developmental and neopatrimonial environments in relation to agents. This study identified islands of high-quality practice in specialised areas of a university or in think tanks

that could be utilized for international development evaluation in all case countries. In neopatrimonial settings agents were more willing to be outwardly critical of policy decisions.

Civil society organizations have become suppliers of evaluations that are legitimate within the power networks and have had success in engaging with policy processes in all of the case countries. In Ethiopia, for example, the Poverty Action Network of Ethiopia has conducted evaluations of initiatives to feed into formal processes. In Zambia, the Jesuit Centre for Theological Reflection is an evaluation, research, education and advocacy organization that promotes study and action on issues linked to Christian faith and social justice in Zambia. The Jesuit Centre for Theological Reflection is an important organization, based upon their history and the prominent position of Christianity in the Zambian constitution.

Local supply agents in many instances are more configured for monitoring of policies, and even more of programmes and projects, for example, annual progress reports. With a few exceptions foreign consultants lead evaluation teams, with limited participation of national consultants. Universities in the case countries, in general, have limited evaluative research expertise. All of the main Universities in the case countries were ranked below 1,300 in the global research production rankings (Scimagoir, 2013). However, in all five countries there is a social science capacity (sociologists, economists, political scientists) which could be mobilized for evaluation work linked to research. The Centre for Disease Control is funding a Monitoring & Evaluation Centre of Excellence at the University of Zambia that could be supportive of high quality evaluation.

Think tanks have received increasing funding in a number of case countries. New think tanks are emerging in Rwanda, Zambia, Ghana, and Malawi. The relative strength of these think tanks appears to relate to their ability to work legitimately within the politics. There are some examples of think tanks that are politically embedded which have helped shift latent demand to actual demand. The African Center for Economic Transformation in Ghana, for example, is an economic policy institute that undertakes policy analysis, evidence-based advocacy and advice to African governments to enable them to formulate and implement good policies and strengthen public institutions towards accelerated development. The African Center for Economic Transformation has undertaken analytical research in areas like foreign direct investment inflows; export promotion policies and strategies; and education and skills development.

CONCLUSION

This chapter illustrates the application of a framework for those who work on international development evaluation processes to identify entry points

to contribute to the improvement of policy by making it more responsive to poor and marginalized people, ultimately contributing to democracy. The principal-agent concepts presented in this paper recognize that undertaking evaluation is inherently an act of power. Being realistic about the constraints of power can help to identify possible entry points for evaluation that applies international development standards.

This chapter has provided a conceptual framework of principals, intermediaries, and agents and identified their roles in development evaluation and presented an emerging understanding of some configurations of power in states that receive development aid. Together, these elements can help identify some potential entry points for evaluation as well as provide initial evidence of the applicability of the principal-agent concepts to implement international development evaluation in a manner that supports responsiveness of governments to poor and marginalized people.

ACKNOWLEDGMENTS

This chapter is based upon work conducted at the Center for Learning on Evaluation and Results for Anglophone Africa under a contract with the Department for International Development and was extensively supported by Salim Latib and Osvaldo Feinstein and David Rider-Smith. The article has benefitted from further guidance on the topic of accountability and responsiveness from Peter Kingstone and Robert Picciotto at the International Development Institute at King's as well as Donna Podems and Charles Amoatey. The author prepared this article in his personal capacity. The opinions expressed in the article are the author's own and do not reflect the view of Oxfam, DFID or any other organization.

NOTES

1. rm evaluation by itself refers to the broader discipline encompassing a range of value positions.
2. For a more detailed discussion on the authorising environment see Benington and Moore (2011)
3. For an elaboration of different types of evaluations used in the U.S. Congress see Chelimsky (2006)
4. The problems with donors operating outside partner country environments was recognized in the Paris Declaration, which sought to work towards five principles of: country ownership; alignment, harmonization, managing for development results, and mutual accountability.
5. The problems with donors operating outside partner country environments was recognized in the Paris Declaration, which sought to work towards five

principles of: country ownership; alignment, harmonization, managing for development results, and mutual accountability.

6. See, for example, Porter (2013) for a discussion of voluntary organizations of professional evaluators as intermediaries.
7. This next section draws extensively on the final synthesis report of the case studies written by Porter and Feinstein, 2014.

REFERENCES

Akerloff, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics, 84*(3), 488–500.

Basinga, P., Gertler, P. J., Binagwaho, A., Soucat, A. L., Sturdy, J., & Vermeersch, C. M. (2011). Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: An impact evaluation. *The Lancet, 377*(9775), 1421–1428.

Bemelmans-Videc, M., Rist, R., & Vedung, E. (2003). *Carrots, sticks, and sermons: Policy instruments and their evaluation*. New Brunswick, NJ: Transaction.

Benington, J., & Moore, M. H. (2011). Public value in complex and changing times. In J. Benington & M. H. Moore (Eds.), *Public value: Theory and practice* (pp. 1–30). London, England: Palgrave Macmillian.

Booth, D., Crook, R., Gyimah-Boadi, E., Killick, T., Luckham, R., & Boateng, N. (2005). What are the drivers of change in Ghana. *CDD/ODI Policy Brief, 1*.

Booth, D., & Golooba-Mutebi, F. (2012). Developmental patrimonialism? The case of Rwanda. *African Affairs, 111*(444), 379–403.

Boyle, R., & Lemarie, D. (1999). *Building effective evaluation capacity: Lessons from practice*. London, England: Transaction Publishers.

Chelimsky, E. (2006). The purpose of evaluation in a democratic society. In I. Shaw, J. C. Green, & M. M. Mark (Eds.), *The Sage handbook of evaluation* (pp. 33–35). Thousand Oaks, CA: Sage.

Dahl, R. A. (1973). *Polyarchy: Participation and opposition.* New Haven and London, CT: Yale.

Diamond, L. J., & Morlino, L. (2005). *Assessing the quality of democracy.* Baltimore, MD: Johns Hopkins University Press.

Fukuyama, F. (2014). *Political order and political decay: From industrial revoloution to globalization of democracy.* New York, NY: Farrar, Straus and Giroux.

House, E., & Howe, K. (1999). *Values in evaluation and social research*. Newbury Park, CA: Sage.

Klerkx, L., & Leeuwis, C. (2008). Matching demand and supply in the agricultural knowledge infrastructure: Experiences with innovation intermediaries. *Food Policy, 33*(3), 260–276.

Lindberg, S. I. (2013). Mapping accountability: Core concept and subtypes. *International Review of Administrative Sciences, 79*(2), 202–226.

Lukes, S. (1974). *Power: A radical view.* London, England: Macmillan.

Mackay, K. (2007). *How to build M&E systems to better support government*. Washington, DC: World Bank.

Mosco, V. (1996). *The political economy of communication: Rethinking and renewal*. Thousand Oaks, CA: Sage

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). London, England: Sage.

Picciotto, R. (1995). Introduction: Evaluation and development. *New Directions for Evaluation, 1995*(67), 13–23.

Picciotto, R., & Wiesner, E. (1998). *Evaluation and development: The instituional dimension*. Washington, DC: World Bank

Porter, S. (2013). Moving beyond teaching people to fish: VOPE's role in sustainable learning strategies by working as innovation intermediaries. In M. Segone & J. Rugh (Eds.), *Strengthening evaluation organization's capacity to contribute to country-led evaluation systems*. New York, NY: UNICEF.

Porter, S., & Goldman, I. (2013). A growing demand for monitoring and evaluation in Africa. *African Evaluation Journal, 1*(1), 9.

Porter, S., & Feinstein, O. N. (2014). *The demand and supply for evaluation in five African Countries*. Retrieved 12 July 2016 from http://www.clear-aa.co.za/wp-content/uploads/2013/09/201402-Demand-and-Supply-Final-Report.pdf

Powell, G. B. (2004). The chain of responsiveness. *Journal of Democracy, 15*(4), 91–105.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Thousand Oaks, CA: Sage Publications.

Schattschneider, E. E. (1960). The Semisovereign People. New York, NY: Holt, Rinehart and Winston.

Schillemans, T., & Busuioc, M. (2015). Predicting public sector accountability: From agency drift to forum drift. *Journal of Public Administration Research and Theory, 25*(1), 191–215.

Scimagoir, (2013). International research rankings. Accessed, November 2013 available at http://www.scimagoir.com/research.php,

Stiglitz, J. E. (2002). Information and the change in the paradigm in economics. *The American Economic Review, 92*(3), 460–501.

Wiesner, E. (2011). The evaluation of macroeconomic institutional arrangements in Latin America. In R. C. Rist, M.-H. Boily, & F. Martin (Eds.), *Influencing change: Evaluation and capacity building driving good practice in development and evaluation* (pp. 23–40). Washington, DC: World Bank.

Appendix F: Porter, S., & Hawkins, P. (2019). ‘Achieving Sustainability through Sustainable Organizational Evaluation Systems’. *New Directions for Evaluation, 2019*(162), 87-101. Available from: https://doi.org/10.1002/ev.20360

Porter, S., & Hawkins, P. (2019). Achieving sustainability through sustainable organizational evaluation systems. In G. Julnes (Ed.), *Evaluating Sustainability: Evaluative Support for Managing Processes in the Public Interest*. *New Directions for Evaluation*, *162*, 87–101.

5

Achieving Sustainability Through Sustainable Organizational Evaluation Systems

Stephen Porter, Penny Hawkins

Abstract

Organizational evaluation systems become sustainable through being networked, use-focused, and by undertaking processes aligned with agreed-upon quality standards. Through being sustainable, organizational evaluation systems can more effectively contribute to the aims and values of sustainable development by undertaking evaluation processes that enable the understanding of complexity, and generate process use, learning loops, and networks of practice. The importance of developing networks of evaluation practice is specifically highlighted in this chapter as it has received less coverage than the importance of use and quality in the literature on evaluation systems. © 2019 Wiley Periodicals, Inc., and the American Evaluation Association.

Introduction

Sustainable development is more likely to be effectively achieved by building sustainable organizational evaluation systems. In this chapter, sustainable development is defined in alignment with the report Our Common Future (World Commission on Environment and Development, 1987, p. 41) which states: "Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs." Evaluation evidence has a role

to play in meeting present needs, without compromising future needs, as Carley and Christie (2000) argue: "Our ability to affect the environment is matched by an inability to assess the consequences of our actions, as we come to realize that natural and human processes are inextricably intertwined" (p. vii). Unsustainable organizational evaluation systems are less likely to contribute to sustainable development as, over a period of time, they will not be able to:

- develop capacity to value changes within complex adaptive systems (Liu et al., 2007);
- develop mechanisms to support the process use of their evidence (Patton, 2016);
- support double loop learning (Argyris, 1977); or
- develop networks to help address complex problems (Ingram, 2015).

The need to think about evaluation systems, their effectiveness, and their contribution toward sustainable development has become pertinent in recent years with the establishment of increasing numbers of organizational evaluation systems. Internationally, the Sustainable Development Goals require country-led evaluation as part of follow-up and review processes (UN, 2015). There has been progress on establishing new evaluation systems within government organizations in a number of countries, such as South Africa and Uganda. In the United States, the Foundations for Evidence-Based Policymaking Act of 2018, establishes in law some key components of evaluation systems, such as evaluation plans and evaluation officers. Evaluation is also becoming a higher priority outside of the government, specifically in the social impact investing field (O'Flynn & Barnett, 2017).

The information presented in this chapter draws upon a range of literature as well as lessons learned from the hands-on experience of the two authors in developing eleven evaluation systems and conducting studies of evaluation systems (see, e.g., Porter & Feinstein, 2014; Porter & Goldman, 2013).

In this chapter, we aim to bring together different strands of writing with the authors' own experiences to elaborate on a kernel of an idea: that organizational evaluation systems, through building three components, become sustainable and can effectively contribute to the aims and values of sustainable development. Sustainable evaluation systems are: (i) networked, (ii) use-focused, and (iii) undertake processes aligned with agreed-upon quality standards. The components of use and quality in evaluation systems have been explored elsewhere, especially in research from the World Bank (Mackay, 2007; Bamberger, 2009; Thomas, 2010; Cunill-Grau &Ospina, 2012; Lopez-Acevedo, Krause, & Mackay, 2012; Routledge, 2018; World Bank, 2018) and the OECD-DAC (OECD, 2016). The network component has received little coverage and is argued here to be important

in developing a sustainable evaluation system. In this chapter, there is an underlying recognition that evaluation systems with all three components (network, use, and quality) can only support sustainable development where the organization values sustainability in its incentives, practices, and culture.

In elaborating on sustainable evaluation systems and their links to sustainable development this chapter discusses: (1) failures in evaluation systems that prevent their sustainability, (2) the three components of a sustainable organizational evaluation system—use and quality, networks. Following these discussions, the authors draw some conclusions.

Failures in Evaluation Systems That Affect Their Sustainability

Unsustainable evaluation systems include those that fail and those that struggle with denuded capabilities, consequently producing lower quality and less useful evaluative evidence. In such situations, evaluation ends up disconnected from decision-making and is not directed toward ongoing improvement to meet current needs. Five issues that affect organizational evaluation systems are discussed in this chapter:

1. an aversion to the reality of results;
2. disfigured accountability;
3. ritualization;
4. empty rhetoric; and
5. censorship.

Aversion to the Reality of Results

Reliable evidence of results can be scary, as such, there can be an *aversion to reality* in organizations. Evidence tells you not only about successes and where success is not being achieved, but also more fundamentally where there is a disjuncture between stated values, assumptions, and the reality of implementation and results. Organizations operate with deeply held assumptions, which can result in them being resistant to evidence (Schein, 2017). For example, in an evaluation managed by one of the authors, an organization believed that it was an excellent partner on social projects. The organization was, however, resistant to evidence contained in an evaluation report that its model of partnership was causing failures in community organizations. This resistance resulted in the organization burying the evaluation results. The rejection of good evidence, in this case, will impact the effectiveness of sustainable development. Community-based development, for example, is recognized to be an element in sustainable development (Carley & Christie, 2000). In not seeking to understand failures in its model of partnership with communities, the organization that buried the evaluation may well cause harm and reduce their ability to support sustainable

development. Without a high-quality evaluation system, evidence is less likely to come to the fore in a manner that helps to work through tensions and disjunctures that reduce the effectiveness of their sustainable development aims.

Disfigured Accountability

Accountability is required for organizational systems to implement strategy in a coordinated manner. Accountability is when a unit knows the nature of the account that is required, how it relates to agreed goals, the person, or body they report to, the kinds of decisions that are going to be taken, and the follow-up process employed (Lindberg, 2013). Accountability processes and evaluation systems work more effectively together, providing knowledge for strategy, when incentives are not overly weighted toward compliance and are supportive of learning and adjustment to needs.

Disfigured accountability, where reporting procedures no longer allow honest reflection and discussion related to organizational values, produces incentives that impede progress toward organizational goals, like sustainability. Accountability in an organization can become disfigured in such a way that evaluation cannot build a focused knowledge base. There can, for example, be silos, which mean evaluations work to serve narrow group agendas and pecuniary interests, rather than the organization's stated goal. Here, the evidence generated through evaluation will not support accountability for organizational strategy, but rather to factions. Second, external accountability demands from others, such as donors and government organizations, can come to displace the accountability for results internally (Boyle & Lemarie, 1999). Here, the evaluation can become a tool of compliance and can result in funded entities just telling others what they want to hear. The authors of this article have seen this happen in non-governmental organizations, where evaluations are produced just for donors and funding organizations. For sustainable development, disfigured accountability can mean that recipient organizations do not take evaluation seriously to improve their performance, while funders receive evidence which is lower quality.

Ritualization

Evaluation can become ritualized. Evaluations are ritualized when they are provided to managers, but they are not taken into account in the actual decision process; instead, weight is given to other sources of information on performance against strategy. Ritualization can result from a belief that evaluation cannot provide quality evidence. If a decision-making body does not find evaluation to be meeting its demands, or questions its methodological rigor, then it will not be taken seriously. As a result, its submission becomes a ritual tick-box exercise. To some extent, evaluators can respond to ritualization by shifting approaches and connecting to the demands of

decision-makers, but only if they are resourced to do so. Ritualization affects the achievement of sustainable development as, again, evidence on effectiveness is side-lined.

Empty Rhetoric

Real champions, who demand evaluations, are required for an evaluation system to function (Kusek & Rist, 2004). Sometimes, however, the rhetoric from ostensible champions on the use of evidence does not lead to the implementation of policies that support evaluation and is known to be a façade for other interests. For example, senior managers may know that a Minister or CEO favors discussions when they are 'evidence-based' (i.e., there is some source of evidence to back them up). In order to build political capital, these managers will start to talk about the importance of evidence and evaluation; however, they will not allocate resources nor actually use evidence in decision-making. The rhetoric is empty. Other managers and staff will know this to be the case. The effect on the evaluation system, is that although there is some profile, there is no real attention to its use and quality; this, in turn, reduces the overall effectiveness of the evaluation system.

Censorship

An organization can have clear strategy, be accountable, and able to recognize failure, but choose to restrict evidence from evaluations through censorship. For a range of reasons, organizations in the public, private, and civil society sectors can be leery about transparency and making performance information publicly available. The authors have observed communication departments, which often have a marketing orientation, focus on presenting a positive version of organizational achievements, for example, to defend the organization against criticism by the media. In working with this positive bias there can be constructive tensions in seeking to present a balanced account of failures and misssteps. There can also be a tendency to suppress information by asking that brief summaries be posted with selected information rather than full evaluation reports.

The kinds of suppression and censorship imposed can range from placing evaluations on separate, hard-to-find websites, to the attempted deletion of whole sections of reports or a refusal to publish reports. These actions whittle down the willingness of managers to undertake evaluation, who with multiple tasks to achieve, may not be willing to put energy into defending evaluations. These sorts of situations can occur when political expediency overrides values relating to openness, honesty, and an ethos of serving the public good (Mathiesen, 2004). These actions will undercut the effectiveness of sustainable development as the flow of information on successes and failures becomes constrained.

All of these examples of failure in an evaluation system will undercut the effectiveness of sustainable development. Organizations without

sustainable evaluation systems can be resistant to evidence of environmental failures, will not hold themselves accountable against the values of sustainable development, and will censor information which shows how they compromise the needs of future generations. When these kinds of challenges mount, evaluation systems become ineffective and produce, for example, episodic evaluation that is not useful, low quality evaluation with a positive bias that is easily disregarded, or practices that are not shared in a network so that the loss of a few key people is debilitating. As a result of these issues, an evaluation system is unlikely to be able to support adaptive management to understand change in complex systems, manage process use, support learning loops, develop networks of practice and so is less likely to change perspectives, policy and programming.

Three Components of Developing Sustainable Organizational Evaluation Systems: Use, Quality, and Networks

Given the amount of writing on what makes for good evaluation, there has been surprisingly little written on what makes an evaluation system sustainable. The main body of writing and experience on defining good evaluation systems and sustaining evaluation systems identified by these authors emanates from the World Bank and the Organization for Economic Cooperation and Development: Development Assistance Committee's (OECD-DAC) Evaluation Network, with important reflections from Chelimsky (2006). The World Bank has published country case studies of evaluation systems (World Bank, 2018), edited volumes on evaluation that incorporate thinking on evaluation systems (Routledge, 2018), journal articles (Cunill-Grau & Ospina, 2012; Thomas, 2010) and guides on building government monitoring and evaluation systems (Mackay, 2007). The OECD-DAC have published studies that map, develop common standards and provide tools for implementing evaluation systems (OECD, 2016). Together this literature provides a good starting point for understanding efforts to sustain evaluation systems, yet even within this literature elements of a good evaluation system are defined sporadically.

Three publications that have definitions and advice on organizational evaluation systems are Mackay (2007), Bamberger (2009), and Lopez-Acevedo, Krause, and Mackay (2012). Between these publications, the components of a successful evaluation system are defined as: a use orientation, good quality evidence, and sustainability. The experience of the chapter authors aligns with the arguments in these texts that relate to evaluation use and quality, with some caveats. In this chapter, sustainability is treated as a higher-level result of an evaluation system that is supported by evaluation use and quality evidence, rather than being a component the same level. The sustainability of the evaluation system only receives limited coverage both conceptually and practically. In existing literature, the sustainability of an evaluation system is argued to result from (i) its use, based on

endogenous demand, and (ii) the quality supplied (Lopez-Acevedo et al., 2012; Lahey, 2012). In this chapter, we argue that the use and quality components are necessary for sustainability, but not sufficient. Additionally, having networked practice is important to support a sustainable organizational evaluation system. Each of the components (use, quality, and networked practice) and their importance in working with and supporting sustainability is outlined in the following pages.

Use

The first component of a sustainable evaluation system is that its products and processes are useful. The idea that good evaluation is useful and applied to help improvement is well documented in the evaluation literature (Shadish, Cook, & Leviton, 1991; Patton, 2008). This idea also extends to evaluation systems, where Lopez-Acevedo et al. (2012) state that a good system is expected to produce information that is "used intensively at one or more stages of the policy cycle" (p. 5). An evaluation system that is useful and used requires that there is demand for evaluations emanating from organizational leadership and structures so that evaluation use is institutionalized rather than episodic (Boyle & Lemarie, 1999).

Incentives and established organizational practices are needed to maintain current, and elicit new, demands for evaluation. Incentives, which are built into policies, procedures, and norms, are argued by Mackay (2007), Vedung (2003), and Boyle and Lemaire (1999), to be especially important in generating demands for evaluation and improving use by making evaluation less scary, more routine, and with defined channels that feed into accountability mechanisms. Mackay (2007) suggests that it is important to have a mix of incentives to generate demand for evaluation; this mix of incentives is described in terms of carrots, sticks, and sermons. Carrots encourage the organization to undertake evaluation, such as, budget allocations being linked to evaluation use. Sticks, such as a withdrawal of funding, penalize people for not undertaking or using evaluation. Sermons extol the importance of evaluation, and occur through actions such as senior leaders talking about how they use evaluations.

Incentivizing new demands is important as it will help evaluation to survive changes in champions, leadership and political administration, whether ministers, CEOs or senior civil service leadership. For example, Lopez-Acevedo et al. (2012) argue that evaluation being "firmly embedded in core government processes, such as, budget cycles" (p. 27) ensures that managers need to consider the use of evaluation when justifying their resources. Put differently, evaluation becomes more sustained when carrots, sticks, and sermons are used to incentivize demand for evaluation evidence through budget processes. This to some extent can be seen in Canada where evaluation is mandatory and has had a consistent regulatory framework put in place over a number of years that enable connection with budgets

(Government of Canada, 2016). To strengthen the demand for evidence on sustainable development through evaluation, the designers of the evaluation system would need to understand organizationally where demand is situated, how it connects to organizational values and assumptions, and build processes and incentives to respond to and strengthen that demand. The advice from these texts on the importance of demand and its renewal through incentives aligns with the experience of the authors of this article. Policies and procedures that favor evaluation, which are built into organizational norms, provide consistency and transparency in expectations and reinforce good accountability practice.

In developing the use component of the evaluation system, it is important that processes are able to perform a challenge function. Evaluations can meet quality standards, connect to internal demands, be independent, and be used in alignment with pre-existing assumptions to make marginal changes. Sustainable evaluation also needs to challenge underlying assumptions around what is being delivered by the organization, and whether the espoused values of the organization connect with its objectives. For example, if an organization is focused on working with marginalized groups of people as partners, it should be possible for evaluations to identify and elaborate where partnership is weak or causing harm. For an evaluation system to be useful it needs to work within accountability systems and be able to hold an organization to account against stated values linked to sustainability.

Quality

The second component of a sustainable evaluation system is that processes and products must be of good quality. This point is also well documented in evaluation practice. A wide variety of texts have made arguments about what constitutes appropriate methodological and process rigor in evaluation from quantitative, qualitative and mixed-method perspectives (Davidson, 2005; Shadish, Cook, & Leviton, 1991; Patton 1990; Banerjee & Duflo, 2017). These standards of appropriate methodological and process rigor are then put into operation for specific organizational environments in developing an evaluation system. Different organizational evaluation systems often define their own typologies of evaluation, interpretations of quality standards, methods, and quality review mechanisms depending on their own institutional demand for evaluation. For example, the evaluation system of the South African Presidency defines six types of evaluation: diagnostic, design, implementation, impact, economic, and synthesis (DPME, 2011). Meanwhile, the U.K. Department for International Development (DFID, 2019) has defined three types of evaluation: process, performance, and impact. Within DFID and the South African Presidency quality is defined as to whether it meets the standards of the type of evaluation being conducted.

Beyond methodological quality, there is also process quality. Elements of process quality include value being placed on transparency, capacity development, management, and accountability. A quality process also promotes use, both while it is happening and through its reporting (Patton, 2010). Further, in developing process quality double loop learning—both detecting implementation issues and questioning underlying goals—can be refined as a part of evaluation practice (Argyris, 1977);. the elements of process quality have been explored elsewhere (Loud & Mayne, 2013; Owen, 2007); here, it is important to note that a sole focus on methodological credibility may limit the use of an evaluation, if for example, it is written in inaccessible language. Process quality helps to build the legitimacy of an evaluation system and the ongoing incorporation of lessons learnt. For example, the experience in the Canadian system indicates that process quality in terms of transparency, capacity, and management are important determinants of the sustainability of an evaluation system (Lahey, 2012).

Although methodological and process quality standards will overlap, an impact evaluation in DFID may apply slightly different criteria of quality than South Africa. It is appropriate that different organizations define their approach to evaluation quality in order that the standards meet with organizational demands, are aligned to incentives, and are therefore more likely to support the use of evaluation.

When evidence of results challenges programmes or managers wish to prevent the communication of findings, people are likely to challenge the quality of the evaluation. Having evaluation quality standards in place helps with mediating arguments and provides a clearer basis from which to judge whether an evaluation should be published or is deemed to have reliable evidence. This helps to facilitate discussions and clarify the boundaries around political and technical concerns about evaluations. For sustainable development, quality standards also help to generate transparency about the range and type of evidence available on evaluations.

Networked Practice

A core idea of this chapter is that evaluation systems that build coordinating, cooperating, and collaborative networks are more likely to be sustainable and contribute toward sustainable development. This is in alignment with literature on sustainability highlighting that networks help address complex problems (Ingram, 2015). Our thinking on networks has developed in response to reflecting on our experience of developing evaluation systems and, in particular, the issues noted earlier of an aversion to the reality of results, disfigured accountability, ritualization, empty rhetoric, and censorship. In this chapter, it is argued that a good evaluation network can function to continually strengthen connections of evaluative thinking and practice, making evaluation less scary and support ongoing reflection and discussion to contribute to adaptive improvement. We recognize that

further empirical work on this component needs to be undertaken and offer reflections based on our experience as well as connections with current evaluation literature.

Camarinha-Matos, Afsarmanesh, Galeano, and Molina (2009) define networking as "communication and information exchange for mutual benefit" (p. 2). They further argue there are more intense and demanding forms of networks including coordination, cooperation and, at the highest end, collaboration. Although information exchange is sufficient for people to know what is happening, more intense forms of collaboration in an evaluation system can help to build demand for evaluation, support adaptive management, and improve quality. Additionally, networks help to resist pressures, such as not discussing results or censoring findings. While it is recognized that networks can have downsides, we argue that a good network function helps sustain evaluation systems. This is because the network can bring together normally autonomous individuals to develop partnerships and capabilities for evaluative thinking, to set common goals on evaluation quality and practice, and to help connect to demand.

In developing the network function of an evaluation system, consideration needs to be given to central, embedded, and external evaluation expertise. As Picciotto and Weisner (1998) highlighted, in an evaluation system there needs to be consideration given to the right mix of independent evaluation, self-evaluation, and what they called outside-in evaluation—other organizations evaluating your performance. The mix of independent, self, and outside-in evaluation is the result of a range of diffuse organizational choices that shape how the evaluation system comes together. These choices are sometimes in the control or near influence of managers of the evaluation system. Working on a network helps to manage the links between independent, self-evaluation, and outside-in evaluation.

Organizations have developed centralized, embedded, or mixed evaluation systems. A centralized evaluation system is mainly built around a single unit, which tends to lean toward more independent evaluations and reporting to a board.[1] A more embedded evaluation system has evaluation (sometimes called monitoring and evaluation [M&E]) staff working more directly with program managers; here independence can be more contested.[2] A mixed system will have elements of both central and embedded.

[1] A central evaluation unit can undertake the following kinds of functions: conduct strategically important evaluations; connect to the demands of leadership across the organization; support the development of organizational performance measurement systems; manage evaluation information systems; set-out and support quality standards for practice in central and embedded elements; develop a use focused ethos for the system; support communication of evaluation findings; help procure good evaluation services; and form networks of people who are mutually interested in evaluation.

[2] The embedded part of an organizational evaluation system may have people who have hybrid roles where evaluation is only part of their work. They have tended to undertake

The choice of system affects the kind of network and their interactions, as well as having consequences for quality and use.

An example illustrates how central and embedded parts of the system affect the development of the networking function. We have worked and observed an organization go through three phases of development in its evaluation system over an 8-year period. In the first phase, a central unit worked mainly to undertake independent evaluations with a close control of to quality and use. In this phase, the evaluation network was limited to mainly information sharing sessions and some external engagements. In the second phase there was a purposeful management directed scale-up of embedded evaluation expertise, with a new arms-length unit to undertake independent review, not necessarily evaluation, activities. Here the central unit ceased to undertake its own evaluations and rather developed mechanisms to support the quality and use of evaluations throughout the system and a network of practitioners. The formal network was defined by professional standards and engagement with an external network of practitioners through funding arrangements. In the third phase, the central unit was reduced in stature and capacity through staff reductions and lowering the organizational rank of the Head of Evaluation. This took place in an organizational context in which there was a greater wariness about presenting results information and a tendency toward inadvertent censorship in an increasingly challenging political context. This reduction in central stature placed a stronger emphasis on network management to continue to sustain professional evaluation practice. These changes in the central and embedded functions reflect the shifting of broader organizational perspectives of evaluation as leaders join and leave the organization and political imperatives change.

In the example above, had there only been a central unit by the time its resources had been reduced, then there would have been substantively less continuing evaluation capability. In practice, a broader network of professional practice and an independent review body continued. In this example, what the network did and continues to do is sustain an evaluation function through three important contributions:

- the development of evaluative thinking capability;
- the definition of and actions to achieve common goals amongst evaluation professionals; and,
- responses to different demands for evaluation evidence throughout the organization.

---

lighter/rapid evaluative activities within the body of the organization like after action reviews; manage evaluations that connect to local management demands; undertake monitoring of performance; facilitate learning discussions; apply standards and use information systems on evaluation.

Evaluative thinking is recognized both at the individual and collective levels. Individually, Buckley, Archibald, Hargraves, and Trochim (2015) argue that:

> Evaluative thinking is critical thinking applied in the context of evaluation, motivated by an attitude of inquisitiveness and a belief in the value of evidence, that involves identifying assumptions, posing thoughtful questions, pursuing deeper understanding through reflection and perspective taking, and informing decisions in preparation for action. (p. 4)

Schwandt (2018) added to this argument that evaluative thinking also has the characteristics of being a deliberative and collaborative process that helps to understand how different interests frame or bound solutions.

Both the individual and group capabilities for evaluative thinking can be furthered through a network of practice. The informal nature of networks allows individuals to test our new ideas, communicate and resolve frustrations, identify more formal training, find existing solutions and access wisdom on how evaluation has worked within an organizational context. More generally, a network can help to get support to overcome issues, such as a resistance to and censorship of results. A formal network enables resourcing for defined interactions and training to encourage the development of evaluative traits.

A network also allows for common goals to be set that reinforce the two other components of an evaluation system. For example, a well-connected network can efficiently help feed into the definition, application, clarification, and disbursement of quality standards for evaluation practice. Additionally, the network provides a broader body of staffing for evaluation steering committees, quality control, and reference groups. When an area of practice needs further development, for example, aggregating quantitative data across organizational silos, a working group can be set-up to help further that practice.

Finally, a network function enables the evolution of the evaluation system to connect to different types of demand for evidence, yielding feedback on what is not working and in doing so develop champions. In diverse organizations, it is likely that there is not just one type of demand. Different actors require evidence for different reasons. Leadership may require strategic evaluations across a portfolio of work, whereas project managers need more rapid feedback loops on what is happening and how things are working. Some of these stakeholders may have a preference for quantitative evidence, while others desire contextual qualitative evidence. A network serves an important function to access these debates, enabling the system, whether central, embedded or mixed, to know what is no longer working, adjust approaches, support resourcing and in doing so, nurture champions of evaluation.

Conclusion

Sustainable development does face many challenges because meeting the needs of the present without compromising the ability of future generations to meet their own needs requires working through trade-offs in different values. Sustainable evaluation systems can help to highlight tensions around such trade-offs and select more effective organizational responses by helping organizations engage in complex interactions and continue to learn through networks. An evaluation system can become sustainable when it improves:

- use, through connecting to and eliciting demand;
- quality, through using common standards on methodology and process;
- networking, through applying a purposeful focus on improving cooperation and collaboration.

These three components help the evaluation system to adapt to challenges as they arise, such as, resistance to the use of evidence, disfigured accountability, ritualization, empty rhetoric, and censorship. An evaluation system will inevitably face new challenges over time, but these three components will be able to adapt and continue to help build evidence of effectiveness within the organization. When the three components use, quality, and network have been informed in their development by the values of sustainable development, then they will, in turn, support the effectiveness of sustainable development.

Acknowledgements

We are grateful to the many colleagues we have worked with who have contributed to our understanding of sustainable evaluation systems. They are too numerous to name here, and we hope they recognise in this article some of the fruitful exchanges had both within organisations and at various professional evaluation events

References

Argyris, C. (1977). Double loop learning in organizations. *Harvard Business Review*, *55*(5), 115–125.

Bamberger, M. (2009). *Institutionalizing impact evaluation within the framework of a monitoring and evaluation system*, Washington, DC: World Bank.

Banerjee, A.V., & ad Duflo, E. (2017). *Handbook of field experiments*. Amsterdam, The Netherlands: North-Holland

Boyle, R., & Lemarie, D. (1999). *Building effective Evaluation capacity: Lessons from practice*, London, England: Transaction.

Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation*, *36*(3), 375–388. https://doi.org/10.1177/1098214015581706

Camarinha-Matos, L. M., Afsarmanesh, H., Galeano, N., Molina, A. (2009) Collaborative networked organizations: Concepts and practice. *Computers & Industrial Engineering*, *57*(1), 46–60.

Carley, M., & Christie, I. (2000). *Managing sustainable development*, New York: Earthscan.

Chelimsky, E. (2006). The purpose of evaluation in a democratic society. In I. Shaw, J. C. Green, & M. M. Mark (Eds.), *The Sage handbook of evaluation* (pp. 33–55). Thousand Oaks, CA: Sage.

Cunill-Grau, N., & Ospina, S. M. (2012), Performance measurement and evaluation systems: Institutionalizing accountability for governmental results in Latin America. *New Directions for Evaluation*, *134*, 77–91. https://doi.org/10.1002/ev.20020

Davidson, J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*, Thousand Oaks, CA: Sage.

Department for International Development (DFID) (2019). Smart Rules: Better Programme Delivery. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/791174/Smart-Rules-External-Apr19.pdf

Department of Performance Monitoring and Evaluation (DPME) (2011). *National evaluation policy framework*. Pretoria, South Africa: South African Government.

Foundations for Evidence-Based Policymaking Act of 2018, 2046 U.S.C. §§312–314 (2018).

Government of Canada (2016). *Policy on Results*. Retrieved from http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=31300

Ingram, J. (2015). Framing niche–regime linkage as adaptation: An analysis of learning and innovation networks for sustainable agriculture across Europe. *Journal of Rural Studies*. *40*, 59–75.https://doi.org/10.1016/j.jrurstud.2015.06.003

Kusek, J. Z., & Rist, R. C. (2004). *Ten steps to a results-based monitoring and evaluation system: A handbook for development practitioners*. Washington, DC: World Bank.

Lahey, R. (2012) The Canadian M&E system. In G. Lopez-Acevedo, P. Krause, & K. Mackay (Eds.), *Building better policies: The nuts and bolts of monitoring and evaluation systems: World Bank training series* (pp. 211–225). Washington, DC: World Bank.

Lindberg, S. I. (2013). Mapping accountability: core concept and subtypes. *International Review of Administrative Sciences*, *79*(2), 202–226.

Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., … Taylor, W. W. (2007). Complexity of coupled human and natural systems. *Science*, *317*(5844), 1513–1516. https://doi.org/10.1126/science.1144004

Lopez-Acevedo, G., Krause, P., & Mackay, K. (Eds.) (2012). *Building better policies: The nuts and bolts of monitoring and evaluation systems: World Bank training series*. Washington, DC: World Bank.

Loud, M. L., & Mayne, J. (Eds.). (2013). *Enhancing evaluation use: Insights from internal evaluation units*. Newbury Park, CA: Sage.

Mackay, K. (2007). *How to build M&E systems to better support government*, Washington, DC: World Bank.

Mathiesen, T. (2004). *Silently silenced: Essays on the creation of acquiescence in modern society*. Winchester, England: Waterside Press.

OECD (2016). *Evaluation systems in development co-operation: 2016 review*. Paris, France: OECD Publishing. https://doi.org/10.1787/9789264262065-en

O'Flynn, P., & Barnett, C. (2017). *Evaluation and impact investing: A review of methodologies to assess social impact*. Retrieved from https://opendocs.ids.ac.uk/opendocs/bitstream/handle/123456789/12835/ER222_EvaluationandImpactInvestingAReviewofMethodologiestoAccessSocialImpact.pdf?sequence=1&isAllowed=y

Owen, J. (2007) *Program evaluation: Forms and approaches*. New York: Guilford.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. CA: Sage.

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.
Patton, M. Q. (2008). *Utilization-focused evaluation*. Thousand Oaks, CA: Sage.
Patton, M. Q. (2016). What is essential in developmental evaluation? On integrity, fidelity, adultery, abstinence, impotence, long-term commitment, integrity, and sensitivity in implementing evaluation models. *American Journal of Evaluation*, *37*(2), 250–265. https://doi.org/10.1177/1098214015626295
Picciotto, R., & Weisner, E. (1998). *Evaluation the institutional dimension*, London, England: Routledge.
Porter, S., & Feinstein, O. N. (2014). *The demand and supply for evaluation in five African countries*, Johannesburg, South Africa: University of the Witwatersrand.
Porter, S., & Goldman, I. (2013). A growing demand for monitoring and evaluation in Africa. *African Evaluation Journal*, *1*, 1. https://doi.org/10.4102/aej.v1i1.25
Routledge (2018). *World Bank series on evaluation and development*. Retrieved from https://www.routledge.com/World-Bank-Series-on-Evaluation-Development/book-series/TRANWBSED
Schein, E. H. (2017). *Organizational culture and leadership*. San Francisco, CA: Jossey-Bass.
Schwandt, T. A. (2018). Evaluative thinking as a collaborative social practice: The case of boundary judgment making. In A. T. Vo & T. Archibald (Eds.), *Evaluative Thinking. New Directions for Evaluation*. *158*, 125–137.
Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.
Thomas, V. (2010). Evaluation systems, ethics, and development evaluation. *American Journal of Evaluation*, *31*(4), 540–548. https://doi.org/10.1177/1098214010373648
UN General Assembly (2015). *Transforming our world: The 2030 agenda for sustainable development*. Retrieved from http://www.refworld.org/docid/57b6e3e44.html
Vedung, E. (2003). Policy instruments: Typologies and theories. In M. L. Bemelmans-Videc, R. C. Rist, & E. Vedung (Eds.), *Carrots, sticks, and sermons: Policy instruments and their evaluation* (pp. 21–58), New Brunswick, Canada: Transaction.
WCED (1987). *Our common future: World Commission on environment and development*. Oxford, England: Oxford University Press.
World Bank (2018). *Evaluators/networks*. Retrieved from http://ieg.worldbankgroup.org/evaluators-networks

STEPHEN PORTER, MPhil, is the director of the Learning, Effectiveness and Accountability Department at Oxfam America and a visiting researcher at the University of the Witwatersrand, Johannesburg. He specializes in Evaluation and Evaluation Systems Development

PENNY HAWKINS is the founder and CEO of an international development evaluation consultancy with offices in New Zealand and Scotland. She works with philanthropic, multilateral, and private sector organizations to align strategy with impact and develop relevant and effective monitoring, evaluation, and learning systems.