Evaluating social accountability interventions: the case for mixed methods and program theory

Sue Cant

Submitted in May 2015 to fulfil requirements of the Master of Evaluation at the Centre for Program Evaluation, Melbourne Graduate School of Education, University of Melbourne
Evaluating social accountability interventions: the case for mixed methods and program theory

Sue Cant, Bachelor of Arts, Master International Business, University of Melbourne. Submitted in May 2015 to fulfil requirements of the Master of Evaluation at the Centre for Program Evaluation, Melbourne Graduate School of Education at the University of Melbourne.

Abstract
Social accountability interventions that promote citizen-state engagement are a relatively recent phenomenon in international development. To date, the evidence of their impact is small, insufficiently robust, but growing. A reasonably strong body of research exists in the international development literature spanning the past decade. However, it is only in recent years that researchers have investigated these interventions from an evaluation perspective. There remains a significant gap in the evaluation literature on guidance for evaluators specific to these interventions and identification of the most suitable evaluation approaches.

The majority of published evaluations of these interventions are conducted through Randomised Control Trials (RCTs). These studies have made a significant contribution to our understanding of the broad components that make up these interventions, including information, collective action and government response. However, to date the qualitative analysis, especially of political context, is lacking, and there are very few high quality evaluations of these interventions using alternative approaches to RCTs. Based on the findings of this paper, theory-based evaluations are advocated drawing on the early evidence from RCTs, using mixed methods and including political analysis through a trans-disciplinary approach.

Declaration
I certify that this is my own original work and due acknowledgement has been made in the text to all other materials used.

Signed: [Signature]
Dated: 15 April 2015
# Table of contents

Acknowledgements .................................................................................................................. 5  
List of acronyms ....................................................................................................................... 6  
List of tables and diagrams ...................................................................................................... 7  
Chapter 1: Introduction ............................................................................................................ 8  
  Overview ................................................................................................................................. 8  
  Background ............................................................................................................................. 9  
  Structure of thesis ................................................................................................................. 11  
Chapter 2: Literature review .................................................................................................... 12  
  Overview ................................................................................................................................. 12  
  Section 1: Phases of the literature review ............................................................................. 12  
  Section 2: Research on social accountability ...................................................................... 13  
  Section 3: Evaluation of social accountability interventions literature ............................. 14  
  Section 4: Evaluation theory, concepts and methods ......................................................... 16  
    Theory-based evaluation ....................................................................................................... 18  
    Recent evaluation approaches and relevant theories ....................................................... 20  
  Section 5: Evaluation of democracy program literature ..................................................... 21  
  Summary .................................................................................................................................. 23  
Chapter 3: Methodology ......................................................................................................... 25  
  Introduction ............................................................................................................................. 25  
  Mixed methods ....................................................................................................................... 25  
  Approach ................................................................................................................................ 26  
    Literature review ............................................................................................................... 26  
    Study 1: Document analysis .............................................................................................. 27  
    Study 2: Consultation with expert panel ............................................................................ 29  
Chapter 4: Results .................................................................................................................... 30  
  Overview .................................................................................................................................. 30  
  Document analysis .................................................................................................................. 30  
    Evaluations against the Joint Committee on Standards for Educational Evaluations  
    (Program Standards) ......................................................................................................... 32  
    Evaluation findings by intervention .................................................................................. 33
Category 1: Information dissemination .................................................................33
Category 2: Community monitoring of services..................................................34
Category 3: Community services scorecards.......................................................35
Category 4: Evaluations using theory of change................................................37
Category 5: Qualitative evaluations ....................................................................40
  Summary of results of document analysis .......................................................41
  Consultation with expert panel .........................................................................41
Chapter 5: Discussion and conclusion ...............................................................48
  Overview ...........................................................................................................48
  The need for qualitative analysis .....................................................................49
  The contribution of RCTs to understanding social accountability interventions ......49
  Theory-based evaluations using mixed methods .............................................50
  Political context ..............................................................................................51
  Donor context ..................................................................................................52
  Recommended approach to evaluation of social accountability .......................52
  Challenges .......................................................................................................53
  Recommendations for research and practice ..................................................54
References ...........................................................................................................55
Appendices ..........................................................................................................64
  Appendix 1 Joint Committee on Standards for Educational Evaluation (Program Standards) .................................................................64
  Utility standards .............................................................................................64
  Feasibility standards .........................................................................................64
  Propriety standards ..........................................................................................64
  Accuracy standards ..........................................................................................65
  Evaluation accountability standards ...............................................................65
  Appendix 2: Social accountability evaluation coding sheet ..............................66
  Appendix 3: List of experts ..............................................................................75
  Appendix 4: Interview questions ......................................................................76
Acknowledgements

Juggling part-time studies with full-time employment places pressure not just on the student, but also on their supervisor and family. I am indebted to Janet Clinton for her intellectual guidance and clarity, and for her grace and interpersonal skills. Janet quickly understood my strengths and weaknesses. She found sensitive and efficient ways to meet our mutual needs, especially considering that she had taken on extra workload to supervise me after the sad passing of Ros Hurworth. I would also like to thank Brad Astbury for his early guidance and encouragement to my thinking.

It is difficult to imagine how I could achieve much in my professional and student life without the unwavering support of my delightful husband, Jim and my family.

I would also like to thank the Melbourne Graduate School of Education’s Centre for Program Evaluation (CPE), from which I am seeking my third degree. The standards maintained by the CPE have been a source of inspiration and pride.
# List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AusAID</td>
<td>Australian Agency for International Development</td>
</tr>
<tr>
<td>CRC</td>
<td>Citizen Report Card</td>
</tr>
<tr>
<td>DFID</td>
<td>Department for International Development</td>
</tr>
<tr>
<td>DFAT</td>
<td>Department for Foreign Affairs and Trade</td>
</tr>
<tr>
<td>GSDRC</td>
<td>Governance and Social Development Resource Centre</td>
</tr>
<tr>
<td>IDRC</td>
<td>International Development Research Centre</td>
</tr>
<tr>
<td>JPAL</td>
<td>Abdul Latif Jameel Poverty Action Lab</td>
</tr>
<tr>
<td>NGO</td>
<td>Non-Government Organisation</td>
</tr>
<tr>
<td>ODI</td>
<td>Overseas Development Institute</td>
</tr>
<tr>
<td>PEA</td>
<td>Political Economy Analysis</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomised Control Trial</td>
</tr>
<tr>
<td>USAID</td>
<td>United States Agency for International Development</td>
</tr>
</tbody>
</table>
List of tables and diagrams

Table 1: Coding criteria
Table 2: Sample of evaluations of social accountability interventions
Table 3: Evaluations by type and number
Table 4: Evaluations by intervention
Table 5: Evaluations of interventions using information dissemination
Table 6: Evaluations of interventions involving community monitoring of services
Table 7: Evaluations of interventions of a community services scorecard
Table 8: Evaluations using a theory of change
Table 9: Qualitative evaluations
Table 10: Themes from expert consultation
Table 11: Triangulation of results against the research questions
Diagram 1: Approach to study
Diagram 2: Model for evaluation of social accountability interventions
Chapter 1: Introduction

Overview

In the wake of the Arab Spring, Robert Zoellick, then head of the World Bank, declared that the international donor community had invested in farms, firms and factories, but not in citizens (Zoellick, 2011). The notion that donor agencies can help citizens engage with their governments for improved accountability has become a central concern of scholars and practitioners of international development over the past decade (Lieberman, Posner and Tsai, 2013). However, despite extensive research to date, there is no definitive consensus on whether such approaches, collectively known as social accountability, work.

Social accountability is the range of actions or strategies, beyond voting, that societal actors, namely citizens, employ to hold the state to account (O’Meally, 2013). These interventions may range from the more conventional, such as donor support for the strengthening of civil society organisations and civic education, to the more innovative, such as participatory budgeting and community monitoring of services. By their nature these interventions are about the relationship between citizens and the state, which make them inherently political, highly contextual and, as they operate within different socio-political and cultural systems, very complex.

The evidence of the impact of social accountability interventions is deemed to be small, insufficiently robust, but growing (Gaventa and Barrett, 2010, Ringold, Holla, Koziol and Srinivasan, 2011, Joshi, 2013a). As a relatively recent phenomenon in international development practice, (World Bank, 2004) the evaluation research in this area is also in its infancy. There are a limited, but increasing number of evaluations of these types of interventions. However, there are very few research studies that discuss evaluation approaches and methods. Several authors have advocated for additional evaluations and more varied evaluation approaches – from both a learning and impact perspective – in order to better assess and inform these interventions (Joshi 2013a, Ringold, Holla, Koziol and Srinivasan, 2011).

In addition, the use of social accountability interventions by multi-lateral, bi-lateral donors and Non-Government Organisations (NGOs) is increasing, irrespective of whether the evidence of their efficacy is well established or not. Essentially, they have been undertaken and evaluated based on normative assumptions and theory, which have not been adequately tested (Gaventa, 2010, Joshi, 2013, Booth, 2012). This gap in the evaluation research has compounded the complexity faced by practitioners in designing, monitoring and evaluating these challenging interventions. Where there is a gap between program implementation and evidence, there is an important role for evaluation research. This study seeks to address this
gap by synthesising the knowledge that exists for those seeking to evaluate social accountability interventions, highlighting the type of evaluations that have been used to date and, consequently, those that may be best suited.

**Background**

**International development and evaluation**

International development programs are social, economic or other programs in developing countries funded by bi-lateral, multi-lateral donors, or other agencies including corporate foundations or Non-Profit/Government Organisations (NGOs). Donors work with a multitude of countries, cultures, governments, actors and agendas. The efficacy of international development efforts is under constant scrutiny (Easterly, 2006, Moyo, 2009, Riddell, 2007, Theroux, 2013). There is often pressure on donors to justify results to domestic political constituencies and long-term development timeframes often conflict with short-term political cycles. Consequently, international development evaluations are driven by the information needs of the major funders (Bamberger, 2000). Until more recently, evaluation in the sector has tended to focus on outputs rather than impact assessment, resulting in weak evidence of impact. In addition, collaboration with the professional evaluation field has been lacking (ibid).

A significant constraint to international aid effectiveness is governance. Corruption, weak accountability and transparency are major impediments to development and poverty reduction and, consequently, are a strong focus of bi-lateral and multi-lateral funders (Ridell, 2007). In order to improve the quality of governance in poor countries, these donors have developed an array of technical programs targeting government bureaucrats and institutions. However, a critical obstacle to institutional strengthening is that of insufficient domestic demand from citizens for better performance or reform by their governments (Teskey, 2005, AusAID, 2007).

There is a growing consensus within the international development community of the significance of the political nature of reform and increasing efforts to ensure that Political Economy Analysis (PEA) is conducted for major donor interventions (Yanguas and Hulme, 2014). An important part of this analysis is the relationship between citizens, state institutions and how these influence reform (AusAID, 2007). The latter has variously been described as democratic governance, demand-led reform, Transparency Accountability Initiatives (TAI) and social accountability, with the latter term adopted here (ibid).

Among more recent approaches trialled to address these governance constraints are social accountability interventions. These approaches involve support to civil society organisations, groups or individuals in activities that seek to highlight the weak performance of government
and, in doing so, place pressure on government for better performance. There is a wide array of activities but their overarching objective is similar: to increase the level of participation by communities in government decision-making and, in doing so, improve the responsiveness and efficacy of governments.

The most prominent of these interventions is participatory budgeting pioneered in Brazil in the 1980s and Citizen Report Cards (CRCs) pioneered in India in the early 1990s. Participatory budgeting involves local people in making decisions on the spending and priorities for a defined public budget. CRCs are reports developed based on household surveys of citizen satisfaction with service delivery, the findings of which have been used by NGOs to advocate to government for improvements. One of the largest programs of its kind in the world, the National Program for Community Empowerment reaching around 77,000 villages in Indonesia, provides an opportunity for communities to decide on where money from government block grants will be allocated (DFAT, 2014).

More recent approaches such as community service scorecards have been introduced by a number of donor institutions and NGOs including the World Bank, CARE, World Vision and Plan. Community scorecards are participatory tools to support communities to develop their own performance and monitoring processes including civic education and advocacy for public services (Ringold, Holla, Koziol and Srinivasan, 2011). The common objective of these approaches is improved service delivery through the increased participation of communities in decision-making on the allocation of public resources.

To help understand how these approaches work, Joshi (2013b) has identified three components: 1) information; 2) collective action; and 3) government response. However, this is a very broad overview, and Joshi’s work (and that of other researchers) is in progress, seeking to delve deeper into the cultural and political behaviours and incentives in order to further articulate triggers for change (Joshi, 2013b, O’Meally, 2013, Tembo 2012). These interventions involve interactions between the state and its citizens, with many actors and relationships involved. It is almost impossible to untangle these relationships from each other, from the political system in which they exist and the cultural norms of that society. These are particularly complex interventions, which despite a substantial body of social and international development research are still not well understood. It is these types of interventions and the quality of their evaluations to date that are the focus of this study.
Structure of thesis

To help understand the nature of these interventions, this study takes a broad view of the research to date.

Chapter 1 provides an introduction and background to the interventions and the rationale for the study's focus. In order to identify the main body of research for these interventions in the international development sector, Chapter 2 scopes international development scholars and the evaluation research of social accountability interventions, of which there are only a handful of main studies. Evaluation theory and concepts are reviewed more broadly. Finally, the most extensive evaluation research in the related thematic area of governance and democracy promotion is examined.

Chapter 3 describes the methodology for the study including the phases of the literature review and the main data collection methods. The results of the data collection from analysis of a sample of 22 evaluations and interviews with an expert panel of evaluators, political scientists and international development academics are presented in Chapter 4. To understand the results in light of the research question and broader environmental constraints faced by evaluators, Chapter 5 is devoted to a combined discussion and conclusion with practical considerations and recommendations presented.
Chapter 2: Literature review

Overview
This study seeks to assess suitable evaluation approaches for donors and practitioners using social accountability interventions. In order to understand the best approaches, a literature review has been undertaken, guided by the following questions:

1. What guidance does the literature provide about evaluation of these interventions?
2. What evaluation approaches have been used to date and what are the gaps?
3. Given the newness of the interventions, might program theory add value in this area?

The literature review is divided into five main sections:

1. An explanation of the phases of the literature review
2. A brief overview of research of social accountability interventions in international development
3. A review of a small number of evaluation studies specifically focused on social accountability interventions
4. A broad review of evaluation theory, concepts, methods and approaches
5. An analysis of the evaluation literature in a related area to social accountability, governance and democracy promotion.

Section 1: Phases of the literature review
Firstly, a brief overview is necessary of the international development research in the area of social accountability. Secondly, the evaluation research on social accountability is surveyed. This is a small number of studies that have identified some of the evaluation approaches used, although little wide-ranging analysis has been made of individual evaluation approaches. However, there is a strong consensus in this literature that theory-based evaluation would be an appropriate evaluation approach.

Thirdly, evaluation theory is reviewed. This literature provides guidance on the origins of the evaluation field, the role of particular research paradigms and their methods including the value of qualitative methods in analysis of context as well as mixed methods in complex interventions. Finally, an examination is undertaken of evaluation research of governance and democracy promotion programs, of which social accountability may be deemed a sub-set due to the similarities between the two areas of practice. This review provides a strong critique of
stand-alone quantitative methods and finds significant gaps in analysis of political context. It suggests overwhelming unanimity that additional theory-based evaluations are required in this area to improve the quality of analysis.

Section 2: Research on social accountability

To understand the history, nature and evaluation of social accountability, it is useful to address the extent of research available in the international development literature.

In 2010, Gaventa and Barrett undertook a meta-case study analysis of a 10-year research program on citizenship, participation and accountability, which analysed the outcomes of four types of citizen engagement approaches through a non-randomised sample of 100 research studies in 20 countries. This study was very broad-ranging but found positive results for the role of citizen participation in four main areas: 1) the construction of citizenship; 2) the strengthening of practices of participation; 3) the strengthening of responsible and accountable states; and 4) the development of inclusive and cohesive societies. Gaventa and Barrett (2010) also found that the outcomes differed depending on the type of citizen engagement and the political context.

At the same time, the Department for International Development (DFID) commissioned a synthesis report of more than 10 years of research in a broader but closely related area, arguing that greater investment was needed in social accountability interventions due to their role in the governance and aid effectiveness agenda (DFID, 2010). In 2011, the World Bank published a book on some of the evidence to date and an in-depth literature review, which highlights the mixed evidence on these interventions and the need for further evaluations in this area (Ringold, Holla, Koziol and Srinivasan, 2011, Bukenya, 2012).

In 2012, David Booth of the Overseas Development Institute (ODI) presented findings of the organisation’s five-year Africa Power and Politics program based on a series of comparative studies, which dedicated a chapter to social accountability interventions. Booth (2012) critiqued the assumptions made about these interventions (which is dealt with further below). He was highly critical of their ability to achieve the governance outcomes suggested in the earlier literature. In addition, a number of other researchers including Joshi, O’Meally, Malena and Tembo have written extensively in this area and, more recently, several have highlighted theories of change for social accountability interventions (Joshi, 2013, O’Meally, 2013, Tembo, 2012).

The extent of research on social accountability over the past 15 years indicates the depth of interest in the international development sector. Indeed, 2013 marked the first dedicated facility to support these interventions, the World Bank’s Global Partnership for Social
Accountability. However, while some of this research has addressed the outcomes of these interventions, there has been limited synthesis of the evidence to date and little research on these interventions from an evaluation perspective. This has been difficult for practitioners trying to design and develop these interventions since the evidence of impact is relatively limited and disparate; and there are still many unknowns about how these interventions work (Joshi, 2013a). More recently, several studies have been published specific to the evaluations of these interventions and it is these studies that we will now examine.

Section 3: Evaluation of social accountability interventions literature

Understanding the thinking behind social accountability interventions is crucial to understanding why and how they have been implemented to date and how to evaluate them. The idea of supporting the demand side of governance – that is, supporting communities to make demands on their governments for better performance – has arisen as a response to political patronage systems in developing countries that have effectively captured votes at elections (Booth, 2012). Interventions supporting communities to demand services between elections have been considered a ‘shorter’ route to accountability than relying only on the electoral cycle (Ringold, Holla, Koziol and Srinivasan, 2011).

Basic services such as health and education are the major responsibility of governments, especially local level governments. A major barrier in government accountability for services is that of asymmetric information between government and poor populations (Lieberman, Posner and Tsai 2013). Principal agent theory posits that one party to a relationship requires a service (the principal or the community) of another party (the agent or the government) but the principal lacks the necessary information to monitor the agent’s performance in any way (Booth, 2012). This theory has been a key driver for social accountability interventions, often resting on assumptions that principals will act on information if they have access to the information. Hence, many of these interventions depend on components that provide information to communities about their governments and services, which is largely not available or accessible to them. However, Booth (2012) is a strong critic of this view, stating that the evidence to support principal agent theory does not exist.

In terms of the role of information more broadly there is a significant amount of evidence. Quite a number of studies have been undertaken on the positive role of information comprising voting behaviour, student enrolment and learning outcomes (Banerjee, 2010a, Chang, 2010). Hence, information is considered a foundational component of social accountability interventions (Joshi, 2013a). However, it is also clear from these studies that information, while a precondition, is insufficient by itself to promote other forms of collective
action (Lieberman, Posner and Tsai, 2013). Thus, it is important for evaluators to be aware that social accountability interventions have been built around principal agent theory and assumptions that information would be sufficient to affect principal agent theory.

In addition to the assumptions around principal agent theory, a number of other factors such as literacy, time and motivation have been assumed of communities involved in these interventions (Beasley and Huillery, 2013). For example, few of these interventions have dealt adequately with the assumption that illiterate communities will feel confident enough to act, that their motivation to act is not inhibited by more urgent economic needs and that when they do act, the agent has both the capacity and the political will to respond (Wild and Harris, 2011).

Joshi (2010) states that due to the lack of specific information on the detail of these interventions and gaps in the evidence, there are many unknowns to consider:

“Despite the popularity of such initiatives, there is little evidence to make emphatic claims about the conditions under which Transparency Accountability Initiatives (TAIs) will lead to effectiveness and impact. The reasons for this are several: vagueness about what an initiative means, the fragmented nature of the evidence, a lack of systematic attention to impact, and few comparative studies that focus on the identification of key enabling factors.” (Joshi, 2010, p 2)

This has significant ramifications for evaluators trying to assess whether these interventions work. If there is no agreed process and outcomes, disparate evidence of how these interventions are working and little documentation on the enabling environment or factors, evaluators are essentially working in the dark.

Social accountability interventions promote citizen state engagement through the use of improved information to the principal (or community) so the principal can hold the agent (or the state) to account for better performance. However, the evidence to date shows that while information is critical, assumptions cannot be made that information automatically creates incentives for principals to exert pressure on agents. These underlying components are important in understanding these interventions in order to evaluate them. However, only a handful of researchers including Shutt and McGee (2012), O’Neill et al. (2007), Roche and Kelly (2012) have considered the features of these interventions from an evaluation perspective.

Shutt and McGee (2012) looked at the evaluability of these interventions on behalf of a number of Non-Government Organisations (NGOs). They identified that most evaluations had fallen into the following categories: experimental, statistical, longitudinal studies, case studies or ethnography, participatory evaluations, meta-analysis or qualitative. However, while the
authors are critical of the quality of monitoring and evaluation undertaken by these NGOs, they do not assess the specific methodologies highlighted and their suitability to these interventions, except in a very broad sense.

Several other authors assessing evaluation of social accountability interventions have argued that theory-based evaluation or theory of change is needed (O’Neill et al., 2007, Roche and Kelly, 2012). O’Neill et al. (2007) state that donor analysis and theory on state and societal transformation, political will and informal politics is weak when these interventions are undertaken.

This section of the literature review has covered evaluation research that is specifically focused on social accountability interventions. Evaluators looking for guidance on the evaluation approaches most suited to these interventions will not find this explicit advice, rather these researchers have a common view on the need for greater political analysis and theory development. Some of the valuable insights they offer include the need to understand that institutions are not neutral and that political will, which can determine outcomes for these interventions, is beyond the control of the intervention. In order to fully explore the guidance in the literature, it is important to understand the main theories, concepts and methods of the evaluation field, which are the subject of the next section.

**Section 4: Evaluation theory, concepts and methods**

Evaluation has been defined as the ‘determination of merit, worth or significance’ (Scriven, 2007). Although evaluation antecedents date to the early 1900s, the evolution of evaluation as a distinct discipline can be traced to the social programs of the Kennedy era (1960s), when evaluation requirements were introduced for federal funding. In the 1970s, when professional evaluation associations emerged, a number of evaluators started to question the dominant role of field experiments and their perceived weaknesses in social programming (Pawson and Tilley, 1997). For example, Guba (1987) and Guba and Lincoln (2001), respectively argued for alternative evaluation approaches such as ‘naturalistic’ or ‘fourth generation evaluation’. Guba and Lincoln (2001) argued that the previous three generations of evaluation were focused respectively on measurement, description and judgement (Carney, 1991).

The fourth generation approach, which may also be described as collaborative inquiry, has suggested a focus on working with and including the interpretation of program stakeholders within the evaluation, something also advocated by Stake (1995) and Chen (1990). Stake (1995) has also argued that there is no one right way to conduct an evaluation, but that evaluations should be ‘responsive’ to the type of intervention under study. Alternatively, Stufflebeam (2007) introduced an evaluation model to ensure that evaluators also considered
the context, inputs, processes and products of an evaluation. These academics, among others, have made significant contributions to evaluation thinking and theories.

However, up to the mid and late 2000s, academics in the field of evaluation remained concerned that epistemological debates still dominated the field. Stame (2004) argued that the question of the black box – the space between the actual input and expected outcome of interventions – still plagued the field of evaluation and was not effectively being addressed. In addition, there is ‘no substantial body of evidence’ on the functioning of different evaluation approaches (Stufflebeam and Shinkfield, 2007, p 189).

As suggested by the lack of evidence on different evaluation approaches in practice, the quest to unpack the ‘black box’ has been affected by the theoretical approach of the evaluation discipline, drawing on historical philosophical paradigms of knowledge generation. Stame (2004) has suggested that this has created a polarising demarcation in the field, which has resulted in a degree of paralysis. Under these competing paradigms, the positivist view advocates that reality exists and can be measured by quantitative methods. In contrast, constructivists argue that reality is created through interaction and negotiation, which can only be conveyed through qualitative methods, though some academics have argued that there is no link between the paradigms and the methods (Guba and Lincoln, 2001). At the same time, pragmatic evaluators in the field have combined qualitative and quantitative methods and argued that evaluations can benefit from multiple paradigms of knowledge generation through the use of multiple methods, which are becoming more popular (Spillane, 2009). Mixed methods allow evaluators to draw on quantitative methods that are appropriate when looking for standardised findings. Nevertheless, to understand a program’s cultural context and dynamics, qualitative methods are considered critical (Stufflebeam and Shinkfield, 2007).

Despite some significant schools of thought on approaches to evaluation, method-driven evaluations and, in particular, RCTs still play a dominating role (Vogt, et al., 2011). This is also true of evaluations in the field of international development, where there has been a similar debate in recent years about the role of RCTs (Ravallion, 2009). Given this methodological influence, it is important to highlight theoretical debates on the use of RCTs to evaluate the type of interventions that are the focus of this study.

To this end, the Governance and Social Development Resource Centre (GSDRC, 2011) has conducted a review of RCT use in social accountability interventions. GSDRC presents arguments by Moehler (2010) suggesting that field experiments can help evaluators to understand universal or generalisable phenomenon. The centre also highlights the position of Bardhan (2011) that RCTs can tell us if an intervention works, but not why or how, which makes it difficult to apply more broadly. In addition, there is persistent doubt about an RCT's
suitability for complex interventions (Jones 2009). Due to the complexity of social accountability interventions and the critiques of the efficacy of RCTs for complex interventions, it is also important to briefly highlight the literature on complex interventions.

An intervention may be considered complex if it has a number of interacting components and variability of outcomes, as occurs with social accountability interventions. O’Neill et al. (2007) argue that social accountability interventions are complex because they are characterised by non-linear and multiple causal chains. Complex interventions pose significant challenges for evaluators, however there is some helpful guidance in the literature. For example, the British Medical Research Council (2008) has produced a guidance statement, suggesting that processes are important yet they should not replace an evaluation on outcomes. Further, it recommends that the implementation of an intervention needs to be comprehensively detailed, while alternatives to RCTs should be considered, and a good theoretical understanding of the intervention is needed. A key recommendation of the council is that mixed methods are appropriate to complex interventions.

This section of the literature has scoped some of the broad thinking in the evaluation discipline, much of which has focused on providing alternative views to experimental designs and method-driven evaluations. Irrespective of the methodological debates and the dominance of method-driven approaches, evaluations require sound evaluation approaches and designs to guide them and the appropriate selection of methods (Joint Committee on Standards for Educational Evaluation, 1994). A number of evaluation frameworks and approaches are supported by sections of the evaluation and international development community that are not driven by methods but by an overarching framework that guides the evaluation. Some of these approaches are selected for examination in the next section, including an influential school of thought that advocates theory-based evaluation, which is discussed next.

**Theory-based evaluation**

Evaluation thinking has evolved over several decades to value and leverage other disciplinary evidence and encourage a more trans-disciplinary approach. Although Stufflebeam (2007) has highlighted a general lack of evidence for particular evaluation approaches, other academics have suggested that evaluators can make use of existing empirical evidence across other disciplines, develop program theories and test these theories using the collection of data against the theory (Chen, 1990, Pawson and Tilley, 1997, Funnell and Rogers, 2011). Theory-based evaluation encourages evaluators to triangulate empirical evidence by drawing on the broadest possible knowledge about program development and social change; be it academic, common sense or first-hand experience (Pawson and Tilley 1997). Rogers (2008) has been an influential advocate of developing a causal modal linking program inputs and
activities to a chain of intended or observed outcomes, and then using this model to guide the evaluation. Rogers (2011) further defines program theory as combining two components: a theory of change and a theory of action. The theory of change details the drivers or processes of change expected or sought out through a program whereas a theory of action highlights how the program is constructed or implemented to activate the theories of change. A theory-based evaluation would analyse and compare the theory of change and the theory of action.

According to Stame (2004), theory-based evaluations have opened up the ‘black box’. They have helped to address paralysis in the field due to the conflict over philosophical positioning on the right way to approach knowledge generation, which was highlighted earlier. Theory-based evaluations challenge method-oriented evaluations and present an opportunity to ensure comprehensive evaluation approaches are considered:

“In this new wave, what changes is the attitude toward methods. There are no more paradigm wars that are immobilizing the field; nor contention about pre-planned multi-method evaluations. All methods can have merit when one puts the theories that can explain a program at the centre of the evaluation design.” (Stame, 2004, p 60)

In advocating theory-based evaluation, Stame (2004) has highlighted the importance of evaluators in critically appraising and designing a sound evaluation based on a reasoned approach that is appropriate to the intervention and not driven by methods.

However, theory-based evaluation is not without its critics. Stufflebeam and Shinkfield (2007) warn that evaluators using this approach may focus their attention on theory developed early in the program and later discover that it has evolved to be a quite different enterprise from what was theorised at the outset.

But others counter that theories do not always have to be right at the outset or uniformly accepted in order to help evaluators better understand the causal chains linking activities, outputs and results. Rather, logic models can be more dynamic and can be reconstructed in order to clarify the linkages between the different dimensions of a program (Weiss 1998).

Theory-based evaluation is useful in international development, given that evidence about the efficacy of specific sectoral development interventions can be sourced in other disciplinary literature, thus utilising the theory-based advocates’ suggestion that evaluators draw on whatever evidence is already available. Theory-based evaluations can also assist considering the fact that interventions in international development tend to be very complex and their efficacy is almost always affected by different cultural contexts. However, until the late 1990s, it appears that evaluators in international development have not looked to theory-based evaluations or other evaluation approaches from the professional evaluation discipline,
perhaps due to a tendency to operate in isolation from the professional and academic evaluation field (Bamberger, 2000).

Indeed, it is worth noting that despite the potential value of approaches such as theory-based evaluations in the international development sector, a number of separate efforts have been exerted over the past decade to improve the quality of evaluation rather than drawing from this evaluation research over some decades. For example, Outcome Mapping is the result of the international development sector’s struggle to deal with attribution to donor interventions in developing countries. This approach, among others, is examined next.

Recent evaluation approaches and relevant theories

It is not within the scope of the study to review all evaluation approaches. However, several approaches and frameworks are selected for their potential relevance to the interventions under study. These are: Developmental Evaluation, Contribution Analysis, and Outcome Mapping. Developmental Evaluation was developed by Patton (2011) to support evaluation of interventions where, as in the case of social accountability, there is no clear model, where the environment is complex and where systems apply.

Patton (2011) has developed some useful questions to guide evaluators, encourage thinking in terms of systems and collect real-time data as part of the ongoing development of a program. Developmental Evaluation is targeted at complex and dynamic interventions like social accountability where the right path or defined outcomes are not necessarily clear or linear (Patton, 2011).

Another approach that may be considered relevant to social accountability interventions is that of Outcome Mapping. It was developed by the International Development Research Centre (IDRC) in the early 2000s to respond to the difficulties in reporting the impact of interventions in diverse cultural contexts, where IDRC argues that attribution is impossible (IDRC, 2001). The focus of this approach is to identify contribution to changes in behaviour, relationships, activities and/or actions of the people, groups and organisations. Social accountability interventions are focused on such changes in behaviour and involve citizen and state relationships as well as collective and individual actions. With Outcome Mapping, outcomes are not pre-determined or assumed. The evaluator identifies the outcomes and works backwards to determine how the program has contributed to the changes that have occurred (ibid). This is especially useful to social accountability interventions where outcomes may be unclear, diverse or not necessarily agreed by all involved.

Similar themes around contribution are found in a further approach, developed by John Mayne. Contribution Analysis assists evaluators to infer causality (Mayne, 2012). It is argued that a reasonable claim of causal contribution can be made if there is a reasoned theory of
change and the key program assumptions are plausible, supported by evidence and agreed upon by at least some of the key players involved in the program.

A number of approaches reviewed here may have relevance to social accountability interventions, since they are highly nuanced in the way they may be implemented and neither the path nor the impact is clearly defined. Evaluators need to consider the most suitable approach to deal with analysis of the complexity of these interventions, including: the importance of context, a good theoretical understanding of the intervention, and the phases of these non-linear interventions. In all likelihood, it is too early to assess the suitability of these more recent approaches and those that may not be considered on their own as sufficient to make conclusions on outcomes. However, many of the analytical frameworks provided by these approaches could be integrated through the consideration of multiple methods.

Multiple methods are implicitly advocated by reviewers of evaluations undertaken in the area of governance and democracy promotion, which is addressed in next section of the literature review.

**Section 5: Evaluation of democracy program literature**

Social accountability interventions involve the promotion of engagement between citizens and the state (O’Meally, 2013). Since citizen-state relations are a crucial part of any democracy, programs that promote democratic institutions and governance are relevant to this study.

A number of program evaluations have been undertaken in the area of democracy promotion and governance. Reviews of these evaluations suggest that evaluators have been too focused on results using quantitative methods, which are unable to capture complex changes involving human political behaviour. The literature also suggests that evaluators have not applied alternative approaches to capture this change. Before detailing the further findings of this section of the literature review, it is useful to highlight the rationale behind interventions, which seek to promote and strengthen democratic institutions and practices, including some of the theory and assumptions that underpin them. The rationale, theory and assumptions as they apply to the topic at hand are important in assessing the relevant evaluation approach.

To review programs relating to governance and democracy promotion, it should be highlighted that the desire on the part of Western nations to spread democratic traditions has a history dating back centuries (Epstein, et al. 2007). Since World War II and earlier, the US has pushed democratic ideals in covert ways (ibid). However, the notion that wealthy countries might support democratisation efforts through development programs in poorer countries as an important, if disputed, pre-condition for stronger economic development and security, has become more prominent over the past two decades (Booth 2012, Collier, 2013).
Along with electoral and leadership support, technical governance programs have emerged to improve accountability in poor and, often, weak states (Booth, 2012). Many of these programs focus on how to improve the accountability of government systems both from the ‘supply’ (or government) side and, more recently, some programs have been established to promote government accountability from the ‘demand’ (or community) side (ibid). The idea that donor support for this demand side might be feasible – where social accountability interventions sit in the governance discourse – is discussed further below.

For the past decade or more, international donors have sought to promote democratic practices and institutions through their overseas aid programs. In the mid 2000s, the Swedish and US Governments undertook meta-studies in relation to the impact of their democracy programs (Finkel, 2006, Uggla, 2007). The US study was unprecedented in its scale, breadth and universal coverage, collecting quantitative data from WWII of all of USAID’s support to democracy programs (Finkel, 2006). Overall, these studies found the impact of these programs to be positive. However, there was criticism of the strong focus on quantitative indicators, the lack of analysis of contextual issues, especially political, and the lack of analysis of the program theory, either through design, or retrospective analysis by the evaluator. There is a strong consensus in the literature about the weakness of evaluative analysis of the causal logic of these interventions and the importance for evaluators to include this in any future evaluations of these types of programs (Finkel, 2006, Uggla 2007, O’Neill et al. 2007, Leeuw, 2003).

Leeuw (2003) argues that evaluations of democracy programs have been unable to shed light on the complexity and context associated with human political behaviour. These include a number of dimensions, which make these interventions challenging to evaluate, including power relations, informal politics and political will. In addition, a number of authors highlight several weaknesses of evaluations in this domain including a focus on results with poor indicators and a need for realistic, measurable, intermediate goals (Burnell, 2007, Uggla 2006, O’Neill et al., 2007).

Evaluations have either been done in an inadequate way or they tend to have been over-engineered, as O’Neill et al. (2007) cite Thomas Carothers: “Democracy promoters have tended either to underdo evaluations, carrying them out hap-hazardly, using superficial methods, or to overdo them, elaborating complex, rigid methods.” (O’Neill et al., 2007, p 32)

However, a key area of consensus emerges on the potential role of theory-based evaluation. Leeuw (2003) argues that most of the problems faced in evaluation of this area are not due to poor resources or implementation, but weak logic. He argues that theory-based evaluations
which focus on the underlying assumptions and rationales are useful in the face of weak evidence from results in democracy promotion.

Burnell (2007) goes further to state that the challenges in these evaluations mean theory-based evaluation is even more crucial: “One could argue that the very features that make support in the area of democracy and human rights difficult to monitor and evaluate on the basis of results; make assessment of the underlying program theory essential for the development of successful programs.” (Burnell, 2007, p 73)

Democracy promotion and governance are areas in which substantial investments have been made by several major international donors. At the same time, these donors have commissioned reviews of the efficacy of these programs through several large-scale evaluations, which have found positive impact. Despite this, some evaluation researchers have taken issue with the methodological weakness of these evaluations including: a predominance of unhelpful quantitative indicators, a gap in the use of program theory and no or insufficient context analysis, in particular political factors pertinent to the program context.

Many that have reviewed these evaluations have advocated the need for use of program theory, theory of change or theory-based evaluation (Leeuw, 2003, Burnell, 2007, O’Neill et al. 2007).

Summary

The general literature review has highlighted key features for evaluators in interventions involving social accountability. Firstly, these interventions are complex and, thus, it is important to be aware of the valuable advice that exists in dealing with complexity. For example: the need for analysis of context, elements of change that may not be linear, that alternatives to RCTs should be considered, and that a strong theoretical understanding of the intervention is important. In the broader evaluation literature there is also an emphasis on consideration of alternatives to RCTs. While a strong evidentiary basis does not exist to unequivocally state which approach is best, there is an influential school of thought in theory-based evaluation that argues for the use of evidence from outside the evaluation discipline.

In addition, there are many clearly reasoned arguments about the need to incorporate qualitative methods due to their ability to highlight context, in particular. The literature also establishes the importance of moving beyond quantitative evaluation approaches, which are deemed inadequate on their own to evaluate these types of interventions. Many of these themes were also identified in the evaluations of democracy and, more specifically, social accountability programs. The strongest and most consistent finding across all the literature is the consensus on the need for more theory-based evaluation in this area. Within this
discussion on the role of theory, it is also important to highlight the need for political analysis as fundamental to understanding these interventions.

For ease of reference, the findings of the literature review are now presented against the sub-research questions.

1. What evaluation approaches have been used to date and what are the gaps?

The literature highlighted that most evaluations are experimental. There are only relatively few qualitative and mixed method studies identified, except those identified by McGee and Shutt (2013). However, these authors were accessing largely unpublished evaluations undertaken by non-government agencies.

2. Given the newness of the interventions, might program theory add value in this area?

There is a clear consensus in the literature that program theory can add value to this area and it is fundamental to evaluation of these types of complex interventions.

Overall, the literature finds that evaluators need to: move beyond sole use of quantitative methods, test the theory of these interventions and ensure that political analysis, non-linearity and context inform the theory of change. While there is a dearth of practical advice for evaluators assessing social accountability interventions, quite recently (2013, 2012) several social accountability experts have raised the need for stronger program theory to help unpack the mechanisms behind social accountability interventions (Shutt and McGee, 2012, Tembo, 2012, O'Meally 2013, Joshi 2013b, Booth 2012).
Chapter 3: Methodology

Introduction
This study seeks to add value to the field of evaluation research by identifying the most suitable evaluation approaches for social accountability interventions and examining the potential role of program theory in improving the quality of these evaluations. The fact that these are new interventions and the evaluation research is limited, are important considerations for the study’s methodology. Complex interventions, such as those considered here, require qualitative analysis. However, since there are so few evaluation studies specific to these interventions, triangulation using different methods is also essential. Therefore, the use of mixed methods is necessary to answer the following research questions:

1. Which evaluation approach might better assist donors and practitioners in the development of social accountability interventions?
   1.1 What guidance does the literature provide about evaluation of these interventions?
   1.2 What evaluation approaches have been used to date and what are the gaps?
   1.3 Given the newness of the interventions, might program theory add value in this area?

Firstly, the rationale for using mixed methods is described. Secondly, the approach and the data collection methods are detailed. These include the literature review and two separate inquiries using document analysis and consultation with an expert panel.

Mixed methods
One rationale for adopting a mixed methods study is its recommended use for complex interventions. Given that substantive issues are inherently complex, as demonstrated under the topic of this inquiry, a mixed methods approach is considered a sensible choice (Greene, 2006, MRC, 2013). The combination of multiple methods also benefits the study through the accommodation of multiple views of knowledge generation.

Like any approach, there are strengths and weaknesses for using mixed methods. For example, combining the methods is likely to be more time consuming and challenging due to the amount of data, the need for integration and the fact that the researcher needs both qualitative and quantitative skills and/or additional researcher/s with these skills. However, as identified by Stufflebeam and Shinkfield (2007), there is a strong consensus that it is a generally appropriate choice.
Approach

It is not within the scope of this study to exhaust the evaluation research and review every evaluation undertaken for these types of interventions. However, to ensure the findings of the literature review are valid, they have been triangulated against two separate inquiries. Firstly, the literature review seeks to establish a strong consensus in theory about some of the evaluation approaches used or advocated for these types of interventions. Secondly, a sample of evaluations was selected for an inquiry using document analysis. Thirdly, this document analysis is used to inform a second inquiry using an expert panel in the area of social accountability interventions.

This approach is depicted in a simple flowchart below:

Diagram 1: Approach to study

Literature review

The literature review has been undertaken using a number of different search engines including standard browsers such as Google, Google Scholar, Firefox and Safari. In addition, the online research databases reviewed include: Sage Research Methods, Web of Science, Web of Knowledge, Social Sciences Research Network and Educational Resource Information Center. Finally, databases of evaluations produced by the major multi-lateral and bi-lateral donors and listed on the website of the Center for Global Development have also been used. These searchable databases include the International Initiative for Impact Evaluation, the Abdul Latif Jameel Poverty Action Lab (JPAL) at MIT University, the World Bank Poverty Evaluations database, the World Bank Independent Evaluation Group’s Impact Evaluations, USAID’s Development Experience Clearing House, the Organisation for Economic Cooperation and Development Assistance Committee Evaluation Resource Center, DFID Evaluation Reports and the Office of Development Effectiveness evaluations database within the Australian Department of Foreign Affairs and Trade. The search terms ‘social accountability’, ‘collective action’, ‘community monitoring’ and ‘community participation in services’ are used separately and in combination with the search term ‘evaluation’.

Firstly, a review of the broad international development literature is undertaken. Secondly, evaluation studies specific to these interventions are examined. Thirdly, the concepts, theories and approaches of the broad evaluation literature are considered. Finally, reviews of
evaluations undertaken of the related area of governance and democracy promotion are scoped. Many of the issues faced by evaluators in this area have relevance to evaluations of social accountability interventions due to their similar context.

**Study 1: Document analysis**

According to the Sage Dictionary of Social Research Methods, document analysis is the detailed examination of a wide range of documents which may include formal and published documents or personal or informal documents including letters and shopping lists (Jupp, V, 2006).

The documents examined in this study are evaluations. The purpose of this document analysis is to identify the most common evaluation approaches used to evaluate social accountability interventions and highlight those that are deemed to have the highest value based on their quality and how they inform practitioner understanding of these interventions.

The 22 evaluations were purposively selected from a pool of 35 evaluations, based on the literature review, key databases and the author’s knowledge gained during her employment related to these specific interventions. Additionally, the evaluations in the sample were selected based on their relevance to the topic under study. The selection criteria for the studies used the nature of the intervention and whether it could be categorised as a ‘social accountability’ intervention or had elements of a social accountability intervention. The key elements, which comprise these interventions are based on Joshi (2013b) core elements for social accountability interventions: 1) information 2) collective action and 3) government response. Joshi (2013b) has been selected since it is one of the first attempts by a researcher to unpack extensive literature on social accountability and develop a program theory for practitioners and evaluators. However, given the relatively small pool of evaluations in this area, evaluations were also chosen if they only contained the element of information. However, in all these interventions information is considered fundamental. The evaluations were selected from a larger pool of evaluations on interventions of the broader area of community participation, decentralisation and service delivery.

To support the analysis, the evaluations are coded by key criteria as detailed below, and their quality is reviewed against the program evaluation standards established by the non-profit standards body, the Joint Committee on Standards for Educational Evaluation, herein referred to as the ‘Joint Committee’.

The standards, later identified in the text and fully detailed in Appendix 1, are based on the five criteria of utility, feasibility, propriety, accuracy and accountability, which address a range of areas including the credibility of the evaluators, their attention to stakeholders and the
relevance of the information in the evaluation. Rather than listing all the standards’ criteria here, the document analysis highlights any issues with the evaluations against the standards and highlights the particular standard in the text. For example, the Joint Committee’s first evaluation standard under Utility is U1: Evaluator Credibility. To assess this standard, the details provided in the evaluation of the evaluator’s professional background, qualifications or employment at a university or research institute are noted. The document analysis uses the following criteria to code the evaluations. A coding sheet has been developed and is included in Appendix 2 of the study. Each evaluation has been coded as depicted in Table 1.

Table 1: Coding criteria

<table>
<thead>
<tr>
<th>Coding criteria</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source and date</td>
<td>Which authors, publication and date?</td>
</tr>
<tr>
<td>Evaluation approach/methods</td>
<td>Is there an evaluation approach and what methods are used?</td>
</tr>
<tr>
<td>Publication</td>
<td>Is it a published or unpublished evaluation?</td>
</tr>
<tr>
<td>Nature of the intervention</td>
<td>Is a detailed description of the intervention provided?</td>
</tr>
<tr>
<td>Social Accountability Theory</td>
<td>Was a theory of change for the intervention identified, used and/or tested</td>
</tr>
<tr>
<td></td>
<td>against the data?</td>
</tr>
<tr>
<td>Quality</td>
<td>Did the evaluation meet the Joint Committee standards?</td>
</tr>
</tbody>
</table>

Using these criteria, the evaluation has been coded based on whether it is published or not, by which authors, date and the journal details. Then, the evaluation approaches and methods are identified as well as the detailed explanation of the intervention or otherwise, whether any prior or developed theory has been used and the data collected against it. Finally, the evaluation is coded based on whether it meets the Joint Committee’s standards.

In addition, further analysis has been undertaken through the categorisation of the evaluations under the nature of the intervention and whether any theory has been used, developed and/or tested. For example, one theory of change by Joshi (2013b) suggests that information and collective action are important components of these interventions. Evaluations have been categorised by intervention against this theory of change – that is, evaluations that involve information provision and/or that promote community collective action have been grouped to note whether the evidence against Joshi’s (2013b) theory is supported.

The evaluations have also been reviewed through the lens of methodological critique. For example, an often-cited issue with RCTs is that of external validity. This refers to the fact that the evidence of an RCT cannot be extrapolated to another setting unless a number of like interventions are undertaken in different contexts. One definition for external validity is ‘the degree to which the conclusions in your study would hold for other persons in other places and at other times’ (Web Center for Social Research Methods, 2014). A number of RCTs of similar interventions have been undertaken in different communities in different countries, so
those with positive results in different locations have been identified. In addition, the evaluator’s understanding of social accountability interventions has been assessed based on the use of theory and whether it has added value to the overall evaluation findings and utility.

**Study 2: Consultation with expert panel**

The second study involves consultation with an expert panel. These experts were identified based on their research or work experience related to social accountability interventions. Since it is such a new area of research, and there are relatively few dedicated evaluation experts in the field of social accountability, it was necessary to look outside the evaluation profession to include social, political and international development researchers. The evaluators and academics have been identified through the literature review. However, expert sampling utilising the professional knowledge of experts in the field of evaluation, social and political research through the author’s employment, have also been used (Palys, 2014).

A list of questions was prepared for one-on-one interviews, which were conducted by phone and email due to the fact that many of the researchers are based overseas. The list of experts and questions is provided in Appendix 3 and 4. The data collected was coded based on themes that emerged from the interviews and identified against the key research questions.

The findings of the literature review have been triangulated against the document analysis and the interviews to produce the findings for the overarching study. The results of the literature review (detailed in Ch. 2), document analysis and the expert consultation are considered in the next chapter.
Chapter 4: Results

Overview

Having reviewed the literature on which evaluation approaches would be useful to these interventions, what gaps exist and whether program theory could add value to this area, the results of data collected to answer these questions are presented based on two separate inquiries utilising document analysis and consultation with an expert panel. Firstly, the results from the document analysis of a selected sample of evaluations of social accountability interventions are detailed. Secondly, the views of an expert panel of evaluators, political science and international development academics are synthesised from one-on-one interviews.

Document Analysis

This section of the results deals with analysis of a sample of 22 evaluations. The evaluations were selected through the literature review and specific databases of evaluations, largely maintained by the major multi-lateral and bi-lateral agencies. Once selected, the evaluations were analysed using coding with the following criteria:

Table 1 Coding Criteria (repeated)

<table>
<thead>
<tr>
<th>Coding criteria</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source and date</td>
<td>Which authors, publication and date?</td>
</tr>
<tr>
<td>Evaluation approach/methods</td>
<td>Is there an evaluation approach and what methods are used?</td>
</tr>
<tr>
<td>Publication</td>
<td>Is it a published or unpublished evaluation?</td>
</tr>
<tr>
<td>Nature of the intervention</td>
<td>Is a detailed description of the intervention provided?</td>
</tr>
<tr>
<td>Social Accountability Theory</td>
<td>Was a theory of change for the intervention identified, used and/or tested against the data?</td>
</tr>
<tr>
<td>Quality</td>
<td>Did the evaluation meet the Joint Committee standards?</td>
</tr>
</tbody>
</table>

The evaluation details have been documented in a coding sheet, presented in Appendix 2, including the approach and methodology used, the details and impact of the intervention and whether the evaluator demonstrated further knowledge of these interventions through use of a theory. The coding included analysis of the evaluations against the Joint Committee program standards: utility, feasibility, propriety, accuracy and accountability. These criteria address a range of areas including the credibility of the evaluator, their attention to stakeholders and the relevance of the information in the evaluation.

Once the raw coding was completed, the evaluations were categorised in groups of interventions for further analysis. However, as there were only two qualitative evaluations these particular evaluations were grouped by evaluation approach. The actual studies, their approaches, methodologies and findings are presented in Table 2.
Table 2: Sample of evaluations of social accountability interventions

<table>
<thead>
<tr>
<th>Evaluation</th>
<th>Approach/Methodology</th>
<th>Finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Adikeshavalu, R, 2004</td>
<td>Qualitative</td>
<td>Positive impact on services</td>
</tr>
<tr>
<td>4. Beasley and Hullery, 2012</td>
<td>Mixed methods (inc. RCT)</td>
<td>Demonstrated parents would participate to improve school outcomes</td>
</tr>
<tr>
<td>5. Bjorkman and Svensson, JPAL, 2009</td>
<td>RCT</td>
<td>Significant impact on child mortality</td>
</tr>
<tr>
<td>6. Chong, A, De La O, A, Karlan, D, &amp; Wantchekon, L, 2013</td>
<td>RCT</td>
<td>Information important but insufficient to improve political accountability</td>
</tr>
<tr>
<td>7. Duflo and Hanna, 2005</td>
<td>RCT</td>
<td>Monitoring teachers reduced absenteeism</td>
</tr>
<tr>
<td>8. Duflo 2007</td>
<td>RCT</td>
<td>Monitoring teachers reduced absenteeism and improved student performance</td>
</tr>
<tr>
<td>9. Ferraz and Finan, 2009</td>
<td>RCT</td>
<td>Information affected voter decisions</td>
</tr>
<tr>
<td>10. Glewwe and Maiga, 2011</td>
<td>RCT</td>
<td>Community action planning led to improved academic performance</td>
</tr>
<tr>
<td>11. Lieberman, Posner, Tsai, 2013</td>
<td>Mixed method</td>
<td>Suggests only minority of citizens take part in civic activism</td>
</tr>
<tr>
<td>12. Olken et al, 2012</td>
<td>RCT</td>
<td>Incentives improve participation and health outcomes</td>
</tr>
<tr>
<td>13. Olken, 2007</td>
<td>RCT</td>
<td>No impact of community monitoring on corruption</td>
</tr>
<tr>
<td>14. Pradhan, 2013</td>
<td>RCT</td>
<td>Improved student learning</td>
</tr>
<tr>
<td>15. Reinikka and Svensson 2004</td>
<td>RCT</td>
<td>Increased enrolment due to information</td>
</tr>
<tr>
<td>21. Wild and Harris, 2011</td>
<td>Qualitative</td>
<td>Improved services and official response</td>
</tr>
</tbody>
</table>

The evaluations are numbered by type of approach/methodology in Table 3.
Table 3: Evaluations by type and number

<table>
<thead>
<tr>
<th>Evaluation approach/methodology</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative</td>
<td>2</td>
</tr>
<tr>
<td>Randomised control trial</td>
<td>13</td>
</tr>
<tr>
<td>Mixed method</td>
<td>3 (additional 4 by same organisation not counted in this category but counted in total)</td>
</tr>
<tr>
<td>Of these above approaches, how many include analysis of theory of change</td>
<td>3 (not counted in total as they fall within mixed method category)</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
</tr>
</tbody>
</table>

(Note that 22 evaluations were included in the sample, and five of the mixed method unpublished evaluations were undertaken by one organisation. This has effectively only been counted once in the table, so as not to distort the extent to which mixed method is being used in published evaluations).

The results demonstrate the dominance of RCTs with 13 of the 22 evaluations (or 13 of 18 if the method adopted by the same organisation is only counted once) categorised as RCTs. There are only two qualitative evaluations, three mixed method studies and only three of the overall studies utilised a theory of change.

The results of the sample highlight the fact that methods drive the evaluations of these interventions and especially RCTs as the primary method. There are few alternatives to RCTs and outside theory-based evaluations, no use of an evaluation approach. Of these studies, the number that uses qualitative methods, mixed methods or a theory of change is very small. These represent clear gaps in the evaluation research.

**Evaluations against the Joint Committee on Standards for Educational Evaluations (Program Standards)**

Of the total number of evaluations, five of the 22 evaluations do not meet the Joint Committee program standards. One evaluation that does not meet the standards is a qualitative evaluation funded by the World Bank. The analysis of quality of this evaluation is dealt with below. The other four evaluations are unpublished evaluations by an NGO. Although, these evaluations don’t appear to meet the standards and are unpublished, they remain in the sample for the value they bring to the discussion. It is important to note that many of these interventions require an intermediary with the community, such as an NGO. However, often NGO evaluations are not readily accessible, as they are not published. To broaden the sample, several unpublished evaluations from the same NGO – using the same intervention but in different countries – are included. It is important to reinforce that the only published evaluation that failed to meet the standards used only qualitative methods.
Evaluation findings by intervention

The results are now presented by type of intervention. The grouping by intervention was selected for two reasons. Firstly, the interventions represent elements of the broad theory of social accountability, posited by researchers, especially Joshi (2013b). Secondly, when grouped by intervention they allow a comparison of results of similar RCTs but in different contexts. The interventions are categorised as follows:

Table 4: Evaluations by intervention

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Evaluations of interventions that involve information dissemination</td>
</tr>
<tr>
<td>2</td>
<td>Evaluations of interventions that involve community monitoring of services</td>
</tr>
<tr>
<td>3</td>
<td>Unpublished evaluations of interventions that involve community monitoring of services</td>
</tr>
<tr>
<td>4</td>
<td>Evaluations of interventions of information dissemination and community monitoring of community using a theory of change</td>
</tr>
<tr>
<td>5</td>
<td>Qualitative evaluations of citizen report cards and scorecards</td>
</tr>
</tbody>
</table>

Category 1: Information Dissemination

The first set of evaluations deals with interventions involving the impact of information disseminated to communities.

Table 5: Evaluations of interventions using information dissemination

<table>
<thead>
<tr>
<th>Information dissemination</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banerjee et al, 2010</td>
<td>Politician report cards increased voter turnout</td>
</tr>
<tr>
<td>Reinikka and Svensson, 2004</td>
<td>Information on school funds increased enrolment</td>
</tr>
<tr>
<td>Chong, De La O, Karlan and Wantchekon, 2013</td>
<td>Information necessary but by itself insufficient</td>
</tr>
<tr>
<td>Lieberman, Posner and Tsai, 2013</td>
<td>Minority of citizens – with better access to education, media, reading and social networks – more likely to take up civic activism</td>
</tr>
<tr>
<td>Ferraz and Finan, 2009</td>
<td>Information on corrupt politicians reduced their vote</td>
</tr>
<tr>
<td>Pandey, 2008</td>
<td>Information on community management role increased student reading</td>
</tr>
</tbody>
</table>

All of these evaluations, except one, which found it was too early to assess impact, demonstrated that information provided to communities in developing countries makes a difference to their decisions or behaviour. Banerjee et al. (2010) found a report card on the performance of politicians increased voter turnout while information on the timing of school budgets increased enrolment (Reinikka and Svensson, 2004). In another study by Pandey (2008), an information campaign to the community about its state-mandated roles and responsibilities in school management showed a significant and positive impact on reading (14-27 percent). Chong, De La O, Karlan, and Wantchekon, 2013 state that while information is necessary, by itself, it may be insufficient to improve political accountability. That is,
communities will not necessarily act on information alone but need other forms of support to build their confidence to engage with government.

These evaluation findings are useful when grouped together as they show the positive role of information provision on community behaviours related to voter response, enrolment rates and student learning. The results provide evidence to support part of the theory of change for social accountability interventions. The next set of interventions use information but also move beyond information provision to promotion of the active engagement of communities with their services such as schools and health clinics. This collective action is the second part of the theory of change espoused by Joshi (2013b).

**Category 2: Community monitoring of services**

The following evaluations assess interventions that may broadly be categorised together as support to communities to monitor services and represent an approach to collective action. For example, in Duflo et al. (2008), students were given cameras to document when teachers arrived in class. In Zeitlin (2011), communities were facilitated to produce scorecards of school performance.

**Table 6: Evaluations of interventions involving community monitoring of services**

<table>
<thead>
<tr>
<th>Community monitoring</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duflo, Dupas and Kremer, 2012</td>
<td>Improved student performance</td>
</tr>
<tr>
<td>Zeitlin, 2011</td>
<td>Improved student test scores, reduced student and teacher absenteeism</td>
</tr>
<tr>
<td>Bjorkman and Svensson, 2009</td>
<td>Reduced child mortality, increased births at clinics and outpatients</td>
</tr>
<tr>
<td>Wild and Harris, 2011</td>
<td>Improved resources and responsiveness from some public officials</td>
</tr>
<tr>
<td>Pradhan, et al, 2013</td>
<td>Improved student learning</td>
</tr>
<tr>
<td>Glewwe and Maïga, 2011</td>
<td>Improved teacher preparation</td>
</tr>
<tr>
<td>Banerjee, et al, 2010</td>
<td>No impact</td>
</tr>
<tr>
<td>Olken, 2007</td>
<td>Little average impact</td>
</tr>
<tr>
<td>Duflo, et al, 2008</td>
<td>Teacher absentee rate dropped</td>
</tr>
</tbody>
</table>

Two of these evaluations found no impact and the remainder found that community monitoring had positive impact on a range of education indicators such as student and teacher performance.

Although the nature of these interventions might be somewhat different, their objective is the same. They are aimed at actively involving communities in the monitoring of their services in order to improve the performance of those services. The central question for each evaluation is the same: Does community participation – in this case monitoring of services – make a
difference? All of these evaluations are RCTs except Wild and Harris (2011) which is a qualitative evaluation that uses a theory of change and is dealt with further below.

All of these evaluations, except two, (Olken 2007, Banerjee, et al., 2010), find positive results for the use of community monitoring on the performance of services. Banerjee et al. (2010) found that people were unaware of the existence of school committees, and providing information and training to the community had no impact on community involvement in schools. Olken 2007 found no impact on corruption of community monitoring in road corruption.

In a further study, Duflo, Dupas and Kremer (2012) found a result indicating useful information about context. They found that giving power to school committees over contract teachers (rather than civil-service teachers) was a key factor for improving teacher effort and student learning. However, Duflo, Dupas and Kremer (2012) also found a negative impact from the intervention where civil-service teachers responded by working less. Finally, the Olken (2012) study finds successful health outcomes from a participatory decision-making process attached to a block grant transferred directly to communities in Indonesia.

This section of the results has highlighted evaluations of interventions involving community-monitoring of services, finding that most of these studies – all of which are RCTs – show positive results. This is interesting when viewed against the broad theory of change, allowing some confidence that collective action has worked in several instances to improve services. The next section deals with the results of evaluations of similar interventions involving communities in improving their services. However, these evaluations have been undertaken by an NGO and are not published.

**Category 3: Community services scorecards**

The following unpublished evaluations were undertaken of the same intervention involving the use of community scorecards of services by the same organisation, but in different countries. The interventions also used civic education and advocacy to support the engagement of communities with their governments.
Table 7: Evaluations of interventions of a community services scorecard

<table>
<thead>
<tr>
<th>Community scorecard</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>World Vision Uganda, 2013</td>
<td>Adhoc improved services and community participation, contribution ambiguous</td>
</tr>
<tr>
<td>World Vision Papua New Guinea, 2013</td>
<td>Adhoc improved services and community participation, some analysis of contribution</td>
</tr>
<tr>
<td>World Vision Armenia, 2013</td>
<td>Adhoc improved services and community participation, contribution ambiguous</td>
</tr>
<tr>
<td>World Vision Kenya, 2013</td>
<td>Adhoc improved services and community participation, contribution ambiguous</td>
</tr>
<tr>
<td>World Vision Zambia, 2013</td>
<td>More evidence of contribution to improved services</td>
</tr>
</tbody>
</table>

Only one of these evaluations meets the Joint Committee standards. Non-compliance to standards is due to the weak or absent methodology sections, data collection and their lack of transparency in identifying limitations of their findings. No details are provided on the samples used and no rationale is provided about their use of quantitative data collection. In addition, none of the evaluations employ the use of an evaluation question to guide the selection of the data collection methods.

One reason for including these agency-led evaluations, despite their weaker quality as measured against the standards, is that they are evaluations of the same methodology, that has separately been evaluated by an Oxford University led randomised control trial on behalf of the same organisation (Zeitlin, 2011). This RCT is also part of the sample. That is, the methodology of the intervention has been tested in an RCT and evidence established of impact on education outcomes. So there is robust evidence that these interventions work through a randomised control trial, but the further evaluations undertaken are unable to attribute causation. However, these evaluations add value through their qualitative, albeit, largely anecdotal rather than systematic data collection.

For example, the qualitative information in these evaluations provides hints on understanding some of the mechanisms for change under these interventions. In the health sector the relationship between frontline health staff and women is of primary importance in increasing utilisation of health centres and, ultimately, health outcomes. These evaluations also highlight the importance of implementation. If the facilitation of the community was not effectively performed, then it was unlikely the community would feel confident and able to take up its own collective action.

This section has detailed the results of unpublished evaluations of a community services scorecard undertaken by the same organisation. The evaluation results are not very clear and show only anecdotal evidence, which cannot be relied upon. However, some of the qualitative analysis is interesting for evaluators seeking to understand the complexity of these
interventions and the theory of how they may work. Results of the analysis on the theory of how these interventions may work, and evaluations that have taken this into account, are detailed next.

**Category 4: Evaluations using theory of change**

The results of the analysis of this set of evaluations are of particular significance since they address one of the main questions of the study: Would program theory add value to evaluations of social accountability interventions?

These evaluations were selected and separated for the purpose of analysis due to the use of a theory of change. There are only three evaluations where the researchers consider a theory of change for the intervention and compare this theory against their data collection. What the evaluations demonstrate through use of the theory is a much stronger analysis of the intervention and, consequently, a better explanation of the results, and also the limitations or expectations possible of these interventions – since some deeper analysis has been considered of what the intervention is trying to achieve. The evaluations utilising theory of change are as follows:

**Table 8: Evaluations using a theory of change**

<table>
<thead>
<tr>
<th>Community monitoring</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild and Harris, 2011</td>
<td>Theory of change too simplistic but there are some examples of improved response by government and improved services</td>
</tr>
<tr>
<td>Beasley and Huillery, 2013</td>
<td>Parental participation supported teachers but only educated parents monitored teacher attendance</td>
</tr>
<tr>
<td>Lieberman, Posner and Tsai, 2013</td>
<td>Minority interested in civic activism and this is dependent on literacy, access to reading materials and social networks</td>
</tr>
</tbody>
</table>

Wild and Harris (2011) evaluated an intervention which involved the provision of information and the use of scorecards by communities to rate the quality of their local services. They developed the theory of change after reviewing the project documents and interviewing project staff. They found the theory, as explained by staff, to be too simplistic and not able to capture the full range of change which they anticipated may be possible through the intervention.

Wild and Harris (2011) also highlighted significant assumptions upon which the intervention was based. These are similar to the assumptions highlighted earlier in the literature including that citizens want to participate in decision-making and will hold public officials to account for the delivery of public services. That is, that when provided with information on entitlements, citizens will make evidence-based demands of supply-side duty bearers; and that public
officials will seek to respond to citizens’ needs where they have adequate information on those needs.

Wild and Harris (2011) respond to the latter assumption, arguing that, although there were some service improvements from the intervention, the main incentive of service providers was to focus on responding to demands from the central government rather than from citizens due to the political dynamics of Malawian patronage systems. Wild and Harris (2011) highlight a key issue faced by donors and practitioners in employing social accountability interventions – what incentives work to prompt government to respond if community pressure is insufficient? For example:

“Attempts to strengthen the demands of citizens – through the provision of information on perceptions of service delivery – are therefore unlikely to have much traction if there are few incentives for the state to respond or where there is little capacity to respond.” (Wild and Harris 2011, p 21)

The issue of incentives raised by Wild and Harris (2012) identifies a significant gap in the research in social accountability and the broader governance discourse about the nature of incentives for government performance.

In another study that utilised a theory of change, Lieberman, Posner and Tsai (2013), find that factors associated with higher civic activism include level of literacy, access to media, accessible reading materials and membership in social networks. They also found in all cases that only a minority of individuals were reasonably likely to take action. These are all important factors for evaluators to understand and evaluate these interventions.

One of the strongest contributions to understanding evaluation needs in the area of community monitoring of services is that of Beasley and Huillery (2012). Their discussion of external expectations and assumptions about community participation are valuable. They state that these interventions are heavily dependent on the “willingness, ability and resources of the beneficiaries”. They further highlight the issues of time and cost for communities and the relationship between service users and providers:

“It may be costly and time consuming to gather local information and may be very difficult in practice to put pressure on doctors, or nurses to improve service quality. The extent to which beneficiaries will be able to surmount these difficulties is likely to depend on the characteristics of the community and, in particular the dynamics of the relationship between the people who are beneficiaries and the people who are performing the service.” (Beasley and Huillery, 2012, p 3)

Beasley and Huillery (2012) suggest the success of community participation in service delivery is highly dependent on context. In order to evaluate the intervention, they construct a
model to explain why previous evaluations have found differences in the effectiveness of community participation in services. The model looks at the relationship between service providers and users, the motivation of users, the cost and benefits of participation and a number of characteristics in the community including desire for education and its long-term benefits to the family, the difference between ‘formal’ authority and ‘real’ authority, the education and wealth of parents, the cost of participation, language and social ties (ibid, p 16).

Under the model developed by the authors, parental participation may take the form of three types of activities, which have been categorised as ‘supportive’ (parents provide labour and materials), ‘managerial’ (parents take an active role in planning/organising) and ‘oppositional’ (parents demand higher quality and take action if necessary). Parents are encouraged to participate in the management and support of their child’s school through an increase in the amount of school funds under the control of the parents.

Drawing on human capital theory, Beasley and Huillery (2012) assume that education is an investment generating benefits and costs for parents who assume benefits from their children’s earning capacity and direct utility from having an educated child (ibid, p 13). Another key assumption is that communities will respond less to empowerment when the taste for the service is low, the initial capacity or empowerment is low or the cost of participation is large.

The model is tested using data collected from an RCT based on an intervention that increased the resources and training to school management committees. Overall, the model is supported by the data collected. Beasley and Huillery (2012) found that where communities had a high taste for education or a low cost of participating, the community was prompt to take supportive actions and managerial actions had a positive impact in all schools. However, the intervention did not impact whether communities took ‘oppositional’ actions such as supervising teacher attendance, except where parents felt they had ‘authority’ to increase pressure on teachers for improved quality (ibid p 13).

This section of the results has reviewed evaluations and highlighted a theory of change for the intervention and the evaluators’ finding against this theory. The results show far more detailed analysis of these interventions is possible through evaluations that draw on a theory of change. One of these evaluations using a theory of change was a high quality published qualitative evaluation that is also referenced in the next section highlighting qualitative evaluations.
Category 5: Qualitative evaluations

Due to their low number, the remaining evaluations have been separated into their own category by type of method. Since the complexity of these interventions require qualitative methods, it was important to highlight the limited number of evaluations drawing only on qualitative methods and the fact that the only published evaluation in the sample which failed to meet the Joint Committee program standards was an evaluation that only used qualitative methods. However, it is important to note that irrespective of the findings on qualitative evaluations it is clear from the data that qualitative methods are critical to evaluations of these interventions and 2) the limited sample of 2 qualitative evaluations mean that it is difficult to make definitive conclusions about whether stand-alone qualitative evaluations would be sufficient for these interventions. The only conclusions that can be drawn are about the two evaluations in the sample. The two qualitative evaluations in the sample are:

Table 9: Qualitative evaluations

<table>
<thead>
<tr>
<th>Evaluation</th>
<th>Finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ravindra, 2004</td>
<td>Citizen report cards led to improved service delivery</td>
</tr>
<tr>
<td>Wild and Harris, 2011</td>
<td>Citizen scorecards can improve services but political dynamics need to be considered in theories of change as the incentive of providers was to respond to higher level governments rather than citizens</td>
</tr>
</tbody>
</table>

The Ravindra (2004) evaluation assessed whether a Citizen Report Card (CRC), based on two household surveys of citizen satisfaction with services, had a positive impact on services. As a result of the surveys, taken five years apart, a number of recommendations were made to the Bangalore Government to improve the quality of service delivery. The key evaluation findings were that the surveys increased public awareness of services, stimulated citizen groups to demand better services and resulted in the introduction by government of a number of reforms to improve service delivery including a property tax, a government taskforce and the streamlining of government agencies’ internal systems and procedures.

This was the only published evaluation in the sample that did not meet the Joint Committee program standards. Firstly, the evaluator was arguably not qualified as an evaluator and potentially biased since he was the former chief secretary of the Government of Karnataka within the state of Bangalore, where the intervention was undertaken. That is, one of the most senior officials targeted by the intervention (see Appendix 1, Joint Committee Program Standards, Utility Standard U1).

Where contribution is indicated, the evidence is relatively weak and may be considered highly subjective. For example, statements are made that are not substantiated and don’t meet the program accuracy standard that conclusions should be explicitly justified or evaluation reasoning made explicit (see Appendix 1, Joint Committee Program Standards, Accuracy A1).
& A7). The evaluator cites a useful example about the introduction of a property tax to help improve infrastructure and services, and claims the CRC influenced the introduction of the tax. There is nothing to substantiate the evaluator's statement. Moreover, while the evaluation outlines that the government made a decision to implement a number of the CRC's recommendations, there is no discussion about other probable factors that may have influenced the decision beyond the CRC. In addition, the time delay between the end of the intervention and the evaluation was six and 10 years respectively, which may raise questions of accuracy where memory is relied upon for the findings.

The results from the other qualitative evaluation by Wild and Harris (2011) are dealt with in the previous section. However, for the purpose of this category, it is important to note that this evaluation did meet the Joint Committee standards and was a high quality evaluation. This assessment is made based on the fact that evaluative judgements were substantiated using the data collected, which was not the case in the Ravindra (2004) evaluation.

This final section of the document analysis has reviewed the qualitative evaluations of social accountability interventions. There were just two evaluations that used only qualitative methods, one of which met the Joint Committee standards and one of which did not. Although both evaluations find positive results for the impact of social accountability interventions, the weak quality of one of the evaluations raises questions about the stand-alone use of qualitative methods for these types of interventions.

**Summary of results of document analysis**

The results demonstrate that, on the whole, evaluators are not using evaluation approaches, rather that methods are driving evaluations. The document analysis reveals a high number of RCTs and a low number of alternative methods are used in evaluations of these interventions including mixed and qualitative methods, of which there are only three and two studies, respectively. This highlights the significant gap in the use of evaluation approaches outside of RCTs. Only one published evaluation failed to meet the program standards of the Joint Committee and it used qualitative methods. The results also demonstrate that mixed method studies are the highest quality when combined with a theory of change.

The next section deals with the second inquiry of the thesis, the consultation with an expert panel.

**Consultation with expert panel**

The second inquiry to support the literature review and the document analysis is a consultation with an expert panel of evaluators, political science and international development academics with knowledge of social accountability interventions. Only two of these experts count social accountability as a dedicated research area. Since there are
relatively few researchers with this experience, it was necessary to draw from a wider cross-
disciplinary section of academics. The use of an expert panel was a valuable data collection
method for triangulation purposes, due to the fact that these interventions are relatively new
and that the evaluation research is so limited.

A number of themes emerged from the interviews. These themes offer some useful insights
into the reasons why evaluation research in this area may be limited to date. Numerous
factors raised point to a constrained environment for evaluators in the area of social
accountability including the limitations of time and budget by multi-lateral and bi-lateral
donors. The complexity of the interventions and, therefore the evaluations, were also
considered a significant issue. Moreover, it was implied that the combination of these factors,
the complexity, the lack of consideration by donors to an appropriate evaluation approach and
the time needed, create a lack of incentive for evaluators to consider these interventions. A
number of other themes were raised and they are represented in Table 10.

Table 10: Themes from expert consultation

<table>
<thead>
<tr>
<th>Expert panel themes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timeframes</td>
<td>Donors allow insufficient time for these types of evaluations</td>
</tr>
<tr>
<td>Evaluation approach/scope/terms of reference</td>
<td>There is insufficient consideration of the types of evaluation approaches that might suit the intervention and, thus, a narrow scope within which evaluators are forced to work</td>
</tr>
<tr>
<td>Complexity</td>
<td>Social accountability interventions are highly complex</td>
</tr>
<tr>
<td>Evaluation skills</td>
<td>Evaluators need a high degree of skill to evaluate these interventions</td>
</tr>
<tr>
<td>Political context</td>
<td>Evaluators need to understand and describe political context as part of the evaluation</td>
</tr>
<tr>
<td>Politics of evaluations</td>
<td>Donors will allow time and budget for RCTs but such time and budget is not necessarily considered for qualitative or mixed method studies</td>
</tr>
<tr>
<td>Lack of agreed outcomes</td>
<td>Donors themselves are unclear about what they understand and expect of these interventions and the evaluation objective/question is unclear</td>
</tr>
<tr>
<td>Constrained donor environment</td>
<td>Overall, these issues create a constrained environment that is not conducive to quality evaluations or appealing to evaluators</td>
</tr>
</tbody>
</table>

The results of the interviews, including these themes, are presented under the broad
headings based on the key questions of the study below:

1. What existing guidance is available to evaluators of these interventions?
2. What approaches have been used and what gaps remain?
3. What role exists for program theory in these types of interventions?
**Existing guidance for evaluators**

The consensus among the expert panel is that the evaluation of social accountability interventions is a particularly challenging area for evaluators. It is challenging due to a number of factors. Firstly, the complexity of the interventions and lack of agreement on their outcomes was raised by a number of those interviewed. Secondly, given this complexity, it was considered that a high level of evaluation skill was required. Thirdly, the average timeframe, which was largely determined by the donor's lack of consideration of the approach needed, was very short. Several experts argued that evaluations of this nature require much longer timeframes, which donors either don’t have or don’t consider. The panel felt that short-term, summative evaluation funding dictated by major donors mean that these interventions are not generally evaluated in a way that is appropriate or fit-for-purpose due to the non-linear, political and long-term change anticipated.

An expert suggested that a range of methodologies requiring a high level of skill was needed:

“This type of evaluation is hard. It requires skills across a range of methodologies and commitment to measuring in the long-term. It is not a short-term evaluation. It requires a much longer timeframe.”

Another expert pointed to the complexity of these evaluations, the role of politics and the lack of incentive for evaluators:

“These are relatively young interventions and they are difficult to evaluate because they are non-linear. It’s not appealing to do studies in this area. It’s too complex, diverse, involves multi-stakeholders, it is messy, non-linear (sic). If you are not evaluating the politics you are doing a bad job.”

The constrained donor environment was highlighted:

“In my experience, operational evaluations frequently face some common challenges including: limited scope, short timeframe and a narrow focus on donor-oriented priorities, and this is also true of many social accountability evaluations.”

Another important consideration was the lack of agreement on the aims of these interventions:

“There are a lot of reasons evaluation is hard to do in social accountability. A lot of social accountability work has different types of goals (and no consensus as to what those goals are), whereas standard development has very clear outcomes. It’s hard to find an outcome that everyone agrees on.”

The expert panel considered that there was a range of factors that had made social accountability a fraught area for evaluation, which limited the available guidance. Overall, the donor environment could be said to be the overarching constraint due to the unrealistic time
frames placed on evaluators to deliver evaluations, with little consideration to the type of evaluation approach that would be appropriate and the time needed to achieve this. Moreover, the complexity of these interventions demands a high level of skill and, coupled with the constrained environment, these factors may be a strong disincentive to evaluators.

**Approaches used to date and the gaps**

The second question to the panel involved the type of evaluation approach that was appropriate, what they had seen, used or read about and any gaps they had identified. Many in the expert panel were not able to say which evaluation approach had been the most commonly used to date, yet most estimated that in terms of methods used there would be a dominance of RCTs. Less were confident of how many qualitative evaluations had been undertaken and their quality. Yet, those who had reviewed qualitative work stated that what was available was relatively poor and inadequate. Several experts made reference to the inappropriate use of RCTs in social programming for complex interventions and their inability to highlight context. There was a general concern that the quality of the evaluations in this area was poor:

"Most evaluations are bad, very few are rigorously formulated, the questions are unclear, where they are clear, the methods are inappropriate. Some have clear outcomes, but not how and why. The general standard is not good."

"The standard of the evaluations was low."

"More of the evaluations are reviews which are useful and interesting, but not substantial enough."

"Qualitative approaches have not been robust enough."

More broadly, there was a strong view that these interventions require an iterative approach – that is, more strategies that might be construed as action research, focus on documenting the process of the intervention and comparative research, including well selected in-depth case studies. For example, more emphasis was suggested on learning strategies and the understanding of these interventions as part of local political processes:

"Recommendations include: more process documentation, more emphasis (and dedicated resources) on learning strategies, more comparative (country-level and regional-level) research, more analysis of social accountability conditions and influencing factors, longer-term impact evaluations, better coordination and sharing of information among donors, more emphasis on social accountability as a political process and set of relationships (versus a set of ‘tools’). A lot of good work is already underway, but there is much to be done."

Another expert highlighted the importance of iterative learning as critical to theory-building in this area:
“You have to work with what you have in an iterative manner. I don’t think it is possible to come up with a grand theory of change. It’s like DNA – you come up with small pieces. It needs an iterative approach – that’s the right way to go for theory building and practice.”

A further expert suggested that politics was a defining feature of these interventions which requires an appropriate approach:

“Methodologies have to respect the nature of the intervention, dynamic, non-linear. An approach is needed that makes sense of this. It is about respecting the features (or) parameters of the object of study. Because of the politics you have to have a narrative of the personalities and the forcefield in which these interventions play out.”

There was an overwhelming consensus that there was a gap in the use of mixed method evaluations available. Overall, the strong consensus was that more mixed method evaluations would produce higher quality results and meet the utility needs of practitioners:

“The first wave of evaluations were great, they got people talking and made people realise this is very complex. The next phase needs to be more mixed methods. It needs process tracing, other qualitative methods. Mixed methods are the best case scenario but can be expensive.”

“It requires multi-methods across a long period of time.”

“It is hard to say if it has been under or over evaluated because it is new. It has either been RCTs or not very good qualitative studies. I haven’t seen any mixed methods.”

In responding to which approaches were important there was strong agreement about the need for qualitative methods in this area and a sense from many experts that multiple methods were needed to address evaluation of social accountability.

**The role of program theory**

Almost all interviewed advocated the use of stronger theory of change analysis as part of any evaluation, be it, social accountability or otherwise. But many said this area, in particular, warranted stronger testing of the theory of change. One of the political scientists interviewed, had read extensively on theory-based evaluation and was a strong advocate for this approach to be used to evaluate social accountability interventions. The RCT experts interviewed were also strong advocates for RCT designs to include theories of change. Of the questions asked of the panel, the role of program theory was considered to be the most important consideration among evaluation approaches. The panel considered that the richness of the data provided through qualitative methods was needed in order to develop theories of change. It was argued that all evaluations should only be done on the basis of a well-reasoned, evidence-based theory of change. It was highlighted that while some evidence
exists about some interventions in some contexts, the understanding of how and why these interventions worked was weak. Without exception, all experts agreed that more emphasis on theory-based evaluation was needed.

“Part of what we are trying to do is build towards the theory of how things work. Qualitative methods tell us about the processes that are difficult to do through RCTs. We can learn about and develop theories of mechanisms from the richness of the data.”

“There has not been enough work done on the theory of change. There is a need for a strong dose of theory-based evaluation.”

“There is some evidence that certain kinds of interventions have certain kinds of impact in certain kinds of contexts, but the evidence about why is weak. You are better to do evaluations about how and why, testing both how and why, and not just the outcome.”

“The best evaluations have started with a theory of change. Evaluation should ideally be done and follow the theory of change.”

“You can’t really do it (evaluation) well without it (ToC).”

“You need to be very explicit in any evaluation of your theory of change.”

There were clear themes arising from the interviews. Firstly, these are very difficult evaluations to undertake. For example, they require ‘iterative’ thinking and approaches that accommodate this kind of thinking. Within this, it was highlighted that the environment for evaluators was also challenging as the donor preference was for summative evaluations. These were deemed by several interviewed to be an inappropriate fit for such interventions. Therefore, it was implicit that evaluators were operating in a somewhat constrained environment when selecting evaluation approaches.

Secondly, there was an overwhelming view that theory-based evaluations were needed to help inform practitioners and donors about these types of interventions. Finally, despite the nature of these evaluations needing ‘iterative’ approaches, while there were arguments for increased use of qualitative methods, not one person interviewed dismissed the role of quantitative methods. Despite, an explicit question about which approach might assist, no expert made a case for stand-alone qualitative evaluations.
When triangulated, the results are quite clear. From the sample, there is a pre-dominance of RCTs, a view also supported by the majority of those interviewed. The most insightful evaluations in the sample are those that discussed a theory of change and used the findings to test the theory or a model. At the same time, the most consistent finding of the interviewees was that theory-based evaluations were needed. Within the sample there were only two qualitative evaluations, one of which was deemed to meet the Joint Committee standards. While many interviewed argued for a greater need for the use of qualitative methods, when asked which evaluation approach would be most appropriate, not one expert argued that more stand-alone qualitative evaluations were needed. However, there was a common view that more mixed method evaluations would add value to evaluation of the field of social accountability.

The expert consultation also shed some light on the reasons behind the results. There were several who commented on the constrained donor context, especially for alternative evaluation approaches to RCTs, which take more time than a conventional summative evaluation. Overall, the consultation revealed that evaluators face challenges on a number of fronts due to the complex nature of these interventions and the political environment in which evaluations are undertaken.

Table 11: Triangulation of results against the research questions

<table>
<thead>
<tr>
<th>Research question</th>
<th>Documentary analysis</th>
<th>Expert consultation</th>
</tr>
</thead>
</table>
| 1. Evaluation approaches and gaps | • RCTs (13)  
• Few mixed methods (3)  
• Qualitative (2)  
• Use of methods not approaches | • Dominance of RCTs  
• Criticism of qualitative evaluations  
• Consensus on need for more mixed method studies |
| 2. Role of program theory | The analysis showed that evaluations which used a theory of change and mixed methods were of a much higher quality | Consensus on theory-based evaluation needed |
Chapter 5: Discussion and conclusion

Overview

The objective of this study was to highlight an appropriate evaluation approach and, in particular, the role of theory-based evaluation, for a unique set of interventions. The study sought to answer key questions about the existing evaluation guidance, approaches and gaps in this area by drawing on a literature review, document analysis and an expert panel. The key question this paper seeks to answer is:

Which evaluation approach might better assist donors and practitioners in the development of social accountability interventions?

Social accountability interventions involve citizens and the accountability of their governments, a relationship that requires a deep understanding and analysis of political context. To date, most evaluations have not included this analysis and there remains little guidance as to how general evaluators, without a political science background, might be expected to undertake this specialist analysis. This lack of political analysis is a significant gap highlighted in the following discussion, which also brings together the findings of the study and a number of recommendations aimed to improve evaluation practice in this area.

This chapter will focus on a number of aspects that highlight the need for stronger evidence of evaluation approaches and processes to support understanding of these interventions and their impact. As the literature review of the evaluation research, the document analysis and the expert panel have established, evaluators are using primarily method-driven evaluations and, in particular, RCTs, to evaluate social accountability. It is also clear that current qualitative methods have not been sufficiently robust to confidently provide evidence of the efficacy of these interventions. There is strong consensus that a robust evaluation framework, along with alternative methods to RCTs, is needed to improve evaluation in this area. RCTs have proved important to understanding these interventions and can provide evidence for a more developed theory of change, something that is urged in the literature, demonstrated in the document analysis and advocated by the experts. Overall, it appears that utilising a theory-based approach to evaluation, underpinned by mixed methods may support the development of the evaluation of social accountability.

The rest of the chapter will discuss these conclusions and recommendations in detail, firstly detailing the high need for qualitative analysis.
The need for qualitative analysis

In surveying the literature, there is a frequently recurring theme on the difficulties of capturing and measuring change of these types of non-linear and dynamic interventions, without recourse to qualitative analysis. This is heavily emphasised in the reviews of evaluations of governance and democracy programs, which need to highlight complex contextual factors such as power relations, informal politics and political will.

Although the arguments in the literature about the role and importance of qualitative analysis are consistent and are confirmed through the findings of this study, the evidence for stand-alone qualitative evaluation approaches remains weak. Through triangulation of two studies of document analysis and expert consultation, the data suggests that, in fact, using qualitative analysis alone has been relatively poor to date.

Despite the agreement on the need for qualitative methods, the quality concerns outlined are surely part of the reasoning for the consensus that mixed methods is needed. However, before discussing this further, it is necessary to respond to the debates about RCTs in the context of social accountability interventions, especially as the vast majority of the evaluations of social accountability interventions are RCTs.

The contribution of RCTs to understanding social accountability interventions

As identified in the literature review and in consultation with the expert panel, there are a number of concerns about the role of RCTs in complex social programs and especially their inability to tell us how an intervention worked (Jones 2009, Bardhan, 2011). Others such as Moehler (2010) maintain that external validity is possible. This study suggests that the understanding of the field of social accountability has only progressed so far because of similar RCTs that have been replicated in different contexts. For example, it has been established that information is a critical component for social accountability. Information has had a positive impact on voter response and has influenced parents to enrol their children into school, in addition to influencing learning outcomes. In addition to the RCTs on the role of information, there are a growing number of RCTs that demonstrate positive impact on services from collective action such as the monitoring of the performance of schools and health clinics by communities.

With this body of work, these studies have enabled us to understand some of the key components of these interventions in a way that helps evaluators to confidently talk about an emerging evidence-based theory of change, specifically for community monitoring of services, within future evaluations. It allows for much greater confidence in the further development of a theory of change, where theoretical supposition by governance and political science experts was once the norm.
However, it is also clear from the data that more stand-alone RCTs are insufficient to address the gap in program theory for these interventions. The data demonstrates that the higher quality of the mixed method studies, based on the sample, and the insights they provide through theory building and qualitative analysis – supported through rigorous quantitative data collection – are needed. Although, it bears repeating that it was difficult to make judgements about using only stand-alone qualitative evaluations because of the small number of published qualitative evaluations available to review. The use of theory-based and mixed method studies means that evaluators need to engage in and consider the worth of qualitative analysis. RCTs have their role and when considering the best method, political and donor context is critical.

**Theory-based evaluations using mixed methods**

The most compelling evidence in the study came from the document analysis of the mixed method studies that included and tested a theory of change for the intervention. For example, unlike many studies to date, the Beasley and Huillery (2012) mixed method study drew heavily on the evidence of a number of RCTs of similar interventions (as described above) and other existing theories – including human capital theory – to understand parental motivation for taking collective action in their children’s schools to monitor performance. With the combination of these theories and the RCT evidence to date, they developed a model that considered the nature of the implementation for promoting parental participation. This detailed and insightful discussion is relatively absent from the other evaluations in the sample.

It is also interesting to note that the researchers’ intervention did not have an impact on whether communities took ‘oppositional’ actions such as supervising teacher attendance, except where parents felt they had ‘real’ versus ‘formal’ authority to increase pressure on teachers for improved quality. While the evaluation drew on comprehensive evidence from interventions about parental participation in schools, it did not leverage the theory established through a number of interventions on the role of information. This begs the question as to whether less educated parents might take more oppositional actions if they were provided with information/civic education about school standards, for example, through the intervention. Would this information, as posited in the other RCTs, give them more confidence to monitor teachers in the same way as more educated parents do? It seems that the provision of information, a crucial pre-condition for collective action, was absent from this intervention and its theory of change, which underscores the need for more theory-based evaluations in this area.

Also enlightening were the other two evaluations that drew on theory to suggest aspects of interventions that may or may not work. For example, Lieberman, Posner and Tsai (2013) questioned if it was realistic to assume all citizens want to be more active in lobbying and
holding government to account. The evidence presented by Wild and Harris (2011) suggested that service providers and public officials may not automatically respond. Factors that were also identified by Lieberman, Posner and Tsai (2013) as required conditions to support this theory, included literacy, access to media, reading materials and membership in social networks. When combined, the analysis provided in these three evaluations using a theory of change, can deliver substantial guidance to evaluators seeking to develop, refine, understand or test theories of change for these types of interventions.

The overwhelming consensus of the literature review, expert panel and document analysis strongly favour the adoption of a theory-based evaluation supported by mixed methods in the assessment of social accountability outcomes. However, the addition of contextual analysis of the political environment in which the intervention is situated would add significant value to this approach. This is the next focus of our discussion.

**Political context**

It is an increasing trend for donors to ensure that they are adopting formal processes of Political Economy Analysis (PEA) in international development (Yanguas and Hulme, 2014). Qualitative analysis of political context is one of the major gaps identified in this study. Perhaps, this is the result of having insufficient evaluators with knowledge of political science or political scientists with knowledge of evaluation and a lack of trans-disciplinary collaboration in this area. While there are a number of evaluation experts that draw on their sectoral knowledge in areas such as health, education, agriculture or criminology, this may not be so common in the political science area. It might be argued that the impact of Pawson and Tilley’s work would have been less significant if they did not understand the discipline of criminology so intimately in order to develop their distinct evaluation framework, which draws heavily on theory and a deep analysis of context.

In the same way, it is hoped that some collaboration of the two disciplines of evaluation and politics might add value to this domain. Certainly, several political scientists and social researchers in this area are already delving into evaluation theories and practice to help them understand the mechanisms of these interventions (See Joshi 2012, Booth 2012, Wild and Harris, 2011 and Wild 2012, Tembo, 2012). Arguably, they could be better supported by the best minds in the evaluation field. The lack of innovative high quality political analysis within the published evaluation research is worthy of further research, especially through a trans-disciplinary approach. Why evaluations may not leverage political science expertise and analysis, and its feasibility, is the subject of the next section.
Donor context
Among the reasons why high quality political analysis may lack feasibility are the issues already discussed concerning the constraints faced by evaluators in international development. Time, budget and weak or non-existent data have already been identified as a significant challenge for standard evaluations in the sector. In addition, as the expert consultations have identified, the changes envisaged under these interventions are not suited to the summative evaluations most donors need, often in great haste. What is needed is a more innovative approach to evaluation that uses evaluators at the outset of a design and throughout the intervention, which is advocated under Outcome Mapping and is, increasingly, being explored in a number of larger interventions. Of course, this is likely to increase the evaluation budget, which is at odds with the current lack of investment in evaluation by some donors due to tighter economic conditions. However, if the number of these interventions is increasing and if the data points to continued weak logic and considerable unsubstantiated assumptions about the way in which these interventions ‘should’ work and the complexity of the context within which they operate, then one might ask why more theory-based evaluations in this area have not been undertaken. It may be, that as highlighted, these interventions are too challenging for evaluators to assess within a constrained donor environment that may, more often than not, accommodate only two evaluative options: an expensive RCT from a prestigious institution or a more conventional summative evaluation with tight time limits. Therefore, the environment in which evaluators operate and the incentives required to undertake theory-based, mixed method evaluations warrants further research.

Recommended approach to evaluation of social accountability
The evidence to date is clear. Evaluation approaches are not being considered to assess and assist practitioners and donors in a complex and innovative area of international development practice. Current evaluations in this area are method-driven evaluations, which are unable to capture the complex political analyses required. There is a high need for strong evaluation approaches, which are the best way to ensure the improved quality and use of evaluations in this area, as advocated by the Joint Committee.

Evaluators assessing the merit of social accountability interventions may draw on research and evidence available to develop the most up-to-date, reasoned and evidence-based theory of change. The use of theory-based evaluation in governance and these specific interventions is well supported by the thematic and broader evaluation literature (Burnell, 2007, Leeuw, 2003, O’Neill, 2007, Joshi, 2013b, O’Meally, 2013, Tembo, 2012, Roche and Kelly, 2012, Stame, 2004, Rogers, 2000 and 2008, Weiss 1995 and 1998, Wild and Harris, 2011, Beasley and Huillery, 2012, Lieberman, Posner and Tsai, 2013). Moreover, it is clear that, as one expert put it, to undertake evaluations in this area without political analysis would result in a
bad evaluation. Donors are requiring PEA for their governance interventions and this advice is being informed increasingly by reputable political scientists (Yangua and Hulme, 2014), despite an outstanding need for trans-disciplinary collaboration. Based on these key needs and the available evidence, a strong evaluation approach with a well-defined process and clear steps becomes possible. Evaluators can employ evaluation practice advocated over many years to ensure that stakeholders inform a well-defined and understood description of the program, and use theory-based evaluation principles to draw on the evidence to date outside the evaluation literature to ensure a strong understanding of the aims of social accountability practice. Once the program is well-defined and informed by these two processes of stakeholder engagement and PEA, a strengthened theory of change using the evidence of a number of RCTs, in combination with the PEA, is feasible.

When the theory of change is developed, a strong theory-based evaluation framework can be designed based on the existing evidence in the literature. This may then be informed through the use of mixed methods, which are deemed the most appropriate approach for complex interventions such as social accountability. This process is illustrated in Diagram 2.

**Diagram 2: Model for evaluation of social accountability interventions**

Challenges

As a new area of research, there were a limited, but certainly sufficient, number of relevant research papers for the literature review undertaken. Within the past few years, the number of research papers has increased rapidly. While only a handful of these address evaluation of these specific interventions in a fairly generic manner, they are still useful and support the major issues raised in broader evaluative literature for complex interventions of this nature. A further limitation encountered is that many of the evaluations were RCTs, which meant that data collected from the evaluation sample was based on a small number of mixed method and qualitative studies. However, there was a clear consensus in the literature and among those interviewed about the importance of mixed methods for complex interventions.
It is also important to note that one question was not clear to all interviewees, although it was not considered an issue by the majority (See Appendix, Question 4). Nevertheless, the question drew out some very interesting and useful material. Although, it might have been better to remove the question for methodological clarity, it was kept because it did not affect the results, but prompted interesting discussion.

**Recommendations for research and practice**

In addition to the need for more theory-based evaluations using mixed methods and political analysis, there is a high need for:

1. Improved synthesis and communication of existing evidence. This would make it far easier for both practitioners and evaluators to build on, and put into practice, theories of change for social accountability interventions. While the Abdul Latif Jameel Poverty Action Lab, for example, provides accessible and succinct policy insights, there is still a substantial amount of evidence it could synthesise about contextual factors, especially, in relation to schools and parental behaviour. RCTs are a highly specialised field and, while synthesis of research is not always appealing to researchers, it is a fundamental need, as demonstrated by the rise in the number of systematic reviews. Yet easier accessibility to synthesised evidence for practitioners and evaluators should not always require the time and rigour of a systematic review.

2. This synthesis of the evidence could be used to support an agreed, more nuanced theory of change, which could be broadly disseminated to NGOs and evaluators.

3. Further research is needed using a trans-disciplinary approach between political scientists and evaluators to identify the contribution of these interventions to improved government response.

4. More research is needed on the constrained donor context and the weak collaboration between the international development and evaluation fields.
References


Funnell S & Rogers P 2011, Purposeful Program Theory: Effective Use of Theories of Change and Logic Models, Jossey Bass Whiley, UK.


Glewwe, P & Maïga, E 2011, ‘The Impacts of School Management Reforms in Madagascar: Do the Impacts Vary by Teacher Type?’, Department of Applied Economics, University of
Minnesota, June, US, viewed 23 October 2013,
<http://www.povertyactionlab.org/publication/impacts-school-management-reforms-
madagascar-do-impacts-vary-teacher-type>.


GSDRC 2011, Helpdesk Research Report: RCTs for empowerment and accountability programs, Governance and Social Development Resource Centre, Commissioned by DFID, 1 April, viewed 24 May 2013,
<http://www.gsdrc.org/docs/open/hd756.pdf&ei=f4yCU6XkHczokAX4rlDwBA&usg=AFQjCNHmugCqqiG80sP2YzcvlqhtgBF8AA&sig2=jrBEGhNaF99MKRUiGAVMg&bvm=bv.67720277,d.dGl>.


Joshi, A 2013b, ‘Context Matters: A Causal Chain Approach to Unpacking Social Accountability Interventions, Work in Progress Paper’, Swiss Agency for Development Cooperation’s Democratisation, Decentralisation and Local Governance Network, April, viewed 3 July 2013,


Appendices

Appendix 1: Joint Committee on Standards for Educational Evaluation (Program Standards)

Utility standards

The utility standards are intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs.

- **U1 Evaluator Credibility** – Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.
- **U2 Attention to Stakeholders** – Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation.
- **U3 Negotiated Purposes** – Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders.
- **U4 Explicit Values** – Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments.
- **U5 Relevant Information** – Evaluation information should serve the identified and emergent needs of stakeholders.
- **U6 Meaningful Processes and Products** – Evaluations should construct activities, descriptions, and judgments in ways that encourage participants to rediscover, reinterpret, or revise their understandings and behaviours.
- **U7 Timely and Appropriate Communicating and Reporting** – Evaluations should attend to the continuing information needs of their multiple audiences.
- **U8 Concern for Consequences and Influence** – Evaluations should promote responsible and adaptive use while guarding against unintended negative consequences and misuse.

Feasibility standards

The feasibility standards are intended to increase evaluation effectiveness and efficiency.

- **F1 Project Management** – Evaluations should use effective project management strategies.
- **F2 Practical Procedures** – Evaluation procedures should be practical and responsive to the way the program operates.
- **F3 Contextual Viability** – Evaluations should recognise, monitor, and balance the cultural and political interests and needs of individuals and groups.
- **F4 Resource Use** – Evaluations should use resources effectively and efficiently.

Propriety standards

The propriety standards support what is proper, fair, legal, right and just in evaluations.

- **P1 Responsive and Inclusive Orientation** – Evaluations should be responsive to stakeholders and their communities.
- **P2 Formal Agreements** – Evaluation agreements should be negotiated to make obligations explicit and take into account the needs, expectations, and cultural contexts of clients and other stakeholders.
- **P3 Human Rights and Respect** – Evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and other stakeholders.
- **P4 Clarity and Fairness** – Evaluations should be understandable and fair in addressing stakeholder needs and purposes.
- **P5 Transparency and Disclosure** – Evaluations should provide complete descriptions of findings, limitations, and conclusions to all stakeholders, unless doing so would violate legal and propriety obligations.
- **P6 Conflicts of Interests** – Evaluations should openly and honestly identify and address real or perceived conflicts of interests that may compromise the evaluation.
- **P7 Fiscal Responsibility** – Evaluations should account for all expended resources and comply with sound fiscal procedures and processes.

**Accuracy standards**

The accuracy standards are intended to increase the dependability and truthfulness of evaluation representations, propositions, and findings, especially those that support interpretations and judgments about quality.

- **A1 Justified Conclusions and Decisions** – Evaluation conclusions and decisions should be explicitly justified in the cultures and contexts where they have consequences.
- **A2 Valid Information** – Evaluation information should serve the intended purposes and support valid interpretations.
- **A3 Reliable Information** – Evaluation procedures should yield sufficiently dependable and consistent information for the intended uses.
- **A4 Explicit Program and Context Descriptions** – Evaluations should document programs and their contexts with appropriate detail and scope for the evaluation purposes.
- **A5 Information Management** – Evaluations should employ systematic information collection, review, verification, and storage methods.
- **A6 Sound Designs and Analyses** – Evaluations should employ technically adequate designs and analyses that are appropriate for the evaluation purposes.
- **A7 Explicit Evaluation Reasoning** – Evaluation reasoning leading from information and analyses to findings, interpretations, conclusions, and judgments should be clearly and completely documented.
- **A8 Communication and Reporting** – Evaluation communications should have adequate scope and guard against misconceptions, biases, distortions, and errors.

**Evaluation accountability standards**

The evaluation accountability standards encourage adequate documentation of evaluations and a meta-evaluative perspective focused on improvement and accountability for evaluation processes and products.

- **E1 Evaluation Documentation** – Evaluations should fully document their negotiated purposes and implemented designs, procedures, data, and outcomes.
- **E2 Internal Meta-evaluation** – Evaluators should use these and other applicable standards to examine the accountability of the evaluation design, procedures employed, information collected, and outcomes.
- **E3 External Meta-evaluation** – Program evaluation sponsors, clients, evaluators, and other stakeholders should encourage the conduct of external meta-evaluations using these and other applicable standards.
Appendix 2: Social accountability evaluation coding sheet

<table>
<thead>
<tr>
<th>Source/date</th>
<th>Methodology</th>
<th>Published</th>
<th>Intervention</th>
<th>Impact/results</th>
<th>Evidence for social accountability theory</th>
<th>JC standard met?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adikeshavalu Ravindra, World Bank, 2004</td>
<td>Qualitative</td>
<td>Yes</td>
<td>Citizen Report Card – randomised household survey of 1140 people’s perception of services in Bangalore. (Done twice, five years apart 1994 and 1999)</td>
<td>Improved services and satisfaction of communities, tax reforms introduced, greater transparency, streamlining of government agencies, new staff training program, new website</td>
<td>Challenges in establishing contribution of the Citizen Report Card to the stated impact. (Interviews undertaken 10 years after event). Where contribution is indicated, the evidence is relatively weak and may be considered subjective. Statements are made that are not always substantiated. Where a stronger link is made out, for example, that the CRC ‘influenced’ the introduction of a tax and that interviewees stated that “public pressure does push bureaucrats to perform”, the statement appears unsubstantiated by even a basic verbatim quote from a decision maker. The strongest objective evidence is that many CRC recommendations were implemented by the Bangalore Government</td>
<td>No</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Study Design</td>
<td>Randomised Control Trial</td>
<td>Community Scorecard</td>
<td>Test Results</td>
<td>Limited Understanding</td>
<td>Evidence of Impact</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------</td>
<td>--------------------------</td>
<td>---------------------</td>
<td>--------------</td>
<td>----------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Zeitlin, Andrew, Oxford University, 2011</td>
<td>Randomised control trial</td>
<td>No</td>
<td>Community Scorecard</td>
<td>Test results showed average student would move from the 50th to the 58th percentile. An 8-10% increase in pupil attendance in the treatment communities using the CVA scorecard. A 13% reduction in teacher absenteeism</td>
<td>Limited understanding of the how and why of the intervention, but strong evidence of cause and effect for participatory community scorecards on improved service delivery and development outcomes. Demonstrated that externally derived scorecards have no impact</td>
<td>Yes</td>
</tr>
<tr>
<td>Bjorkman and Svensson, JPAL, 2009</td>
<td>Randomised control trial</td>
<td>Yes</td>
<td>Community Scorecard</td>
<td>33% reduction in child mortality, 58% increase in birth deliveries, 19% increase in outpatient numbers</td>
<td>Strong evidence of impact. Limited understanding of the how and why of the intervention but dramatic evidence of cause and effect in child health outcomes</td>
<td>Yes</td>
</tr>
<tr>
<td>Banerjee, A, 2010, JPAL</td>
<td>Randomised control trial</td>
<td>Yes</td>
<td>3 treatments: 1) Providing information on existing institutions 2) training community members in a testing tool for children 3) training volunteers to hold remedial reading camps</td>
<td>No impact on community involvement, teacher effort, or learning outcomes except improved reading skills</td>
<td>Negative findings on information and community monitoring contradicting many other studies. Study notes context and participatory processes crucial to impact. The participatory approach used and the community’s level engagement with the information may have limited the impact. Description of the intervention quite vague – ‘got discussions going about service delivery’. Distributed</td>
<td>Yes</td>
</tr>
<tr>
<td>Study</td>
<td>Type of Study</td>
<td>Findings</td>
<td>Main Evidence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Banerjee et al, 2010, JPAL</td>
<td>Randomised control study</td>
<td>Yes, Report cards handed out on politicians before election</td>
<td>Higher turnout, reduced vote buying, and higher vote share for better performing incumbents.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Olken, 2007, Oxford University</td>
<td>Randomised control study</td>
<td>Yes, Increased grassroots participation by community members in local-level monitoring of road projects</td>
<td>Little average impact on reducing corruption. Most evidence connected to social accountability’s link to improved services. This suggests no impact for social accountability on reduction of corruption.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beasley et al, 2012, Columbia University</td>
<td>Mixed method</td>
<td>Yes, Increased decision-making and financial control for parents in education</td>
<td>All parents increase participation in ways that support the teachers, but only educated parents increase monitoring of teacher attendance. Tests interesting context model against RCT to show that communities will respond less to community empowerment when the taste for the service is low, the initial capacity or empowerment is low, or the cost of participating is large. The authors found that increasing the financial resources under the control of parents increased participation to support the school and supplement the grant with their own inputs, especially those.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
parents with higher benefit from enrolling their child and lower cost of participating. Parents also engage in activities that help school management. Only educated parents directly oppose teachers

<table>
<thead>
<tr>
<th>Location</th>
<th>Methodology</th>
<th>Social Audit</th>
<th>Social Audit Details</th>
<th>Evaluation Consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>World Vision Zambia, 2013</td>
<td>Mixed method</td>
<td>No</td>
<td>Social audit of government standards, community scorecard, interface meeting with government to develop action plan for improved services 4/5 clinics where intervention used increase staffing</td>
<td>Has strong examples of adhoc results but no consistency of data, no sample, no clear picture of what the evaluation question is that is being answered. Patchy, anecdotal information</td>
</tr>
<tr>
<td>World Vision Uganda, 2013</td>
<td>Mixed method</td>
<td>No</td>
<td>Social audit of government standards, community scorecard, interface meeting with government to develop action plan for improved services</td>
<td>Some examples of anecdotal results, but no consistency off data to answer an evaluation question</td>
</tr>
<tr>
<td>World Vision Kenya, 2013</td>
<td>Mixed method</td>
<td>No</td>
<td>Social audit of government standards, community scorecard, interface meeting with government to develop action plan for improved services</td>
<td>Some examples of anecdotal results, but no consistency off data to answer an evaluation question</td>
</tr>
<tr>
<td>World Vision Armenia, 2013</td>
<td>Mixed method</td>
<td>No</td>
<td>Social audit of government standards,                                                                VISION</td>
<td>Some examples of anecdotal results, but no consistency off data to answer an evaluation question</td>
</tr>
<tr>
<td>Study <strong>World Vision, Papua New Guinea, 2013</strong></td>
<td>Methodology</td>
<td>Social Audit</td>
<td>Community Scorecard, Interface Meeting with Government and Advocacy</td>
<td>Data to answer an evaluation question</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>Duflo, Hanna, 2005 National Bureau of Economic Research</strong></td>
<td>Randomised Control Trial</td>
<td>Yes</td>
<td>No</td>
<td>This was a very useful evaluation because of its discussion of context</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Social Audit of Government Standards, Community Scorecard</td>
<td>Teacher Absentee Rate dropped by an average of 20% (from an average of 42% to 22%)</td>
<td>This is a simple intervention with clear-cut evidence of the fact that monitoring teachers reduces their absentee rate, demonstrating the effect of community monitoring</td>
</tr>
<tr>
<td><strong>Duflo et al 2007, Harvard University</strong></td>
<td>Randomised Control Trial</td>
<td>Yes</td>
<td>Multi-faceted Intervention that combined local hiring of teachers with increased community monitoring</td>
<td>Evidence that monitoring teachers reduces their absentee rate and improves academic performance</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Civil-service teachers – where the school committee was empowered to monitor teachers – were more likely to be in class and teaching during random visits, and their students performed better than those without committee monitoring</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td><strong>Pradhan, et al, 2013</strong></td>
<td>Randomised Control Trial</td>
<td>Yes</td>
<td>Elections of School Committees and Joint Planning Meetings</td>
<td>Found changes were cost effective at improving student learning</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Block Grant Intervention allowing Community</td>
<td>Villages with incentives had improved results on</td>
<td>Evidence that incentives improve the impact of</td>
</tr>
<tr>
<td><strong>Olken, et al 2012</strong></td>
<td>Randomised Control Trial</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Study (Glewwe, P and Maiga E, 2011)</td>
<td>Study Type</td>
<td>Randomised control trial</td>
<td>Description</td>
<td>Outcome 1: Increased District and teacher supervisory training. Distribution of report cards to schools on their previous year’s dropout rate, exam pass rate, and repetition rate. Community action planning based on the report card</td>
</tr>
<tr>
<td>Wild &amp; Harris 2011 et al, Overseas Development Institute</td>
<td>Qualitative</td>
<td>Yes</td>
<td>Community scorecard</td>
<td>Changes in how communities approach local service blockages, shifts in resources and evidence of greater responsiveness from some public officials</td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>Intervention Details</td>
<td>Findings</td>
<td>Notes</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>----------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Reinikka and Svensson, Journal of Public Economics, 2004</td>
<td>Randomised control trial</td>
<td>The intervention was a newspaper campaign in Uganda aimed at reducing corruption of public funds by providing schools (parents) with information to monitor local officials' handling of a large education grant program</td>
<td>Schools with information increased share of entitlements, enrolments increased and achievement improved</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Information campaign led to schools claiming higher share of entitlements. Those schools experienced increased enrolment. Positive effects on achievement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferraz and Finan, 2009, National Bureau of Economic Research</td>
<td>Randomised control trial</td>
<td>Brazilian Government random audit of municipalities prior to municipal elections. Authors compared electoral outcomes of municipalities before and after 2004 elections when audit information was released</td>
<td>Release of the audit outcomes had a significant impact on incumbents’ electoral performance. Effects were more pronounced in municipalities where local radio disseminated information. Findings also suggest that electoral rules enhance political accountability by playing a crucial role in constraining politician’s corrupt behavior</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Findings highlight value of having a more informed electorate and the role of local media in enhancing political selection</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Study Authors</td>
<td>Study Design</td>
<td>Methodology</td>
<td>Result 1</td>
<td>Result 2</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>-------------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>Chong, A, De La O, A, Karlan, D, &amp; Wantchekon, L 2013, et al, 2010 National Bureau of Economics</td>
<td>Randomised control trial</td>
<td>Yes</td>
<td>Corruption and public expenditure information campaign conducted a week before the 2009 local elections in Mexico. Information highlighted mayor's responsibility for the provision of public lighting, safe water, sewage and local roads</td>
<td>Information about corruption decreases incumbent support in elections, decreases voter turnout, challengers' votes, and erodes voters' identification with the party of the corrupt incumbent</td>
</tr>
<tr>
<td>Lieberman, Posner and Tsai (2013) et al 2013</td>
<td>Mixed method</td>
<td>Yes</td>
<td>Information campaigns to promote citizenship</td>
<td>No substantive effect</td>
</tr>
</tbody>
</table>
Appendix 3: List of experts

1. **Chris Roche**, Associate Professor International Development Latrobe University
2. **Dr. Andrew Zeitlin**, Researcher, Georgetown University
3. **Dr. Hans Antlov**, Social Development Adviser, World Bank Jakarta
4. **Dr. Gill Westhorp**, Consultant Evaluator
5. **Dr. Irene Gujit**, Consultant Evaluator
6. **Dr. Jess Dart**, Managing Director, Clear Horizons
7. **Dr. David Booth**, Political Scientist, Overseas Development Institute
8. **Dr. Stephen Kosack**, Researcher, Harvard University
9. **Dr. Anu Joshi**, Researcher, Overseas Development Institute
10. **Courtney Tolmie**, Governance Adviser, Results for Development Institute
11. **Dr. Carmen Malena**, Director, Participatory Governance Program, CIVICUS
12. **Shawn Powers**, Policy Manager, Abdul Latif Jameel Poverty Action Lab
13. **Don Parafina**, Executive Director, Affiliated Network for Social Accountability East Asia
14. **Dr. Simon O'Meally**, Governance Adviser, World Bank, Delhi
Appendix 4: Interview questions

1. Is there a solid evidence base for SA interventions?
2. If so, has the evidence base been established through evaluations or other research?
3. If not, why?
4. If not, how does that effect evaluations of SA interventions?
5. Have there been sufficient evaluations?
6. Generally, what is your opinion of evaluations you have seen of these interventions?
7. What types of evaluations have been more common and why do you think that?
8. Joshi (2013) states that there is a strong normative belief in citizen-led accountability without a clear understanding of the conditions under which it can have impact. Is it good enough to undertake evaluations in these circumstances?
9. If not, how might evaluation research address this situation and what methodologies could be used?
Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Cant, Suzanne

Title:
Evaluating social accountability interventions: the case for mixed methods and program theory

Date:
2014

Persistent Link:
http://hdl.handle.net/11343/54651

File Description:
Evaluating social accountability interventions