International Development Group Working Paper Series

Guide to Assessing Social Accountability Efforts Across Sectors

Taylor Williamson

Guide to Assessing Social Accountability Efforts Across Sectors

Taylor Williamson

Global Health Division,
International Development Group (IDG),
RTI International
701 13th Street, NW, Suite 750
Washington, DC 20005-3967
USA
Tel: 202-974-7889
E-mail: rtwilliamson@rti.org

TABLE OF CONTENTS

ABBREVIATIONS

CCT	conditional cash transfer
CDD	community-driven development
DRG	USAID Centre of Excellence on Democracy, Human Rights, and Governance
EMAS	Expanding Maternal and Neonatal Survival, Indonesia
GAPP	Governance, Accountability, and Participation Program, Uganda
DHS	Demographic and Health Survey
DRG	Center of Excellence on Democracy, Human Rights, and Governance, USAID
IDG	International Development Group, RTI International
MP	Member of Parliament
NGO	nongovernmental organization
RCT	randomized controlled trial
RTI	RTI International (registered trademark and trade name of Research Triangle Institute)
USAID	United States Agency for International Development

ABSTRACT

Since the 2004 *World Development Report* articulated the theoretical linkages between governance and service delivery, the United States Agency for International Development (USAID), the World Bank, and other key donors have promoted embedding social accountability concepts within sector-specific programs. In the interest of leveraging these investments, this paper builds on existing research to pinpoint key concerns for linking social accountability inputs to various types of outcomes. It also describes key concerns for program managers as they try to assess the effectiveness of social accountability programs. I argue that assessing social accountability efforts requires considering a variety of different inputs, contextual factors, study design limitations, and operational concerns. Additionally, the recent emphasis on experimental designs has given program managers more knowledge about what works in specific contexts, but we are often still in the dark about why interventions work. To address this limitation, program managers can use qualitative methods, including causal pathway analysis, to enrich experimental or observational data. Finally, I find that the divide between research that discovers new interventions and programs that implement known "best practices" will only shrink, and program managers must be ready to propose evaluation approaches that respond to the needs of donors and governments seeking to implement social accountability programs.

ACKNOWLEDGMENTS

INTRODUCTION

Since the *World Development Report 2004* articulated the theoretical linkages between governance and service delivery, program designs have integrated technical interventions with governance approaches (World Bank, 2003). Both the United States Agency for International Development's (USAID's) 2012 *Global Health Strategic Framework* and 2013 *Strategy on Democracy, Human Rights and Governance* highlight the interdependence of governance and service delivery challenges and promote embedding governance concepts within sector-specific projects (USAID, 2012, 2013a).

A critical element of governance approaches is social accountability. According to the World Bank (2007, p. 5), social accountability refers to "the broad range of actions and mechanisms, other than voting, that citizens can use to hold the state to account as well as actions on the part of government, civil society, media and other actors that promote or facilitate these efforts." Social accountability interventions often seek to improve one of three "instrumental aims": governance and democracy, citizen empowerment, and service delivery (Brinkerhoff & Wetterberg, 2015).

To continue to build on its investments in social accountability, USAID has emphasized the need to assess the causal links between social accountability inputs and service delivery outcomes. The Center of Excellence on Democracy, Human Rights, and Governance (the DRG Center) has a learning agenda that highlights the need to assess the contribution of governance inputs to service-delivery outputs (USAID, 2014b; Health Finance and Governance, 2014). USAID has also funded research to study social accountability interventions through Development Innovations Ventures grants (USAID, 2014a).

For implementers, programmatic research is critical to ensuring that investments in international development are directed to where they can do the most good. Ensuring that social accountability programs are rigorously assessed, and that outcomes are clear to donors, is also an ongoing concern. This paper is meant to support programs that integrate social accountability interventions by improving knowledge of assessment options at various design stages. The paper offers its target audiences—international donors and technical assistance providers, local nongovernmental organizations (NGOs), and researchers—a consolidated review of the considerations and mechanisms used to assess social accountability interventions in individual country contexts; and it provides details to these audiences on how to apply these mechanisms to future social accountability programs. However, the paper stops short of providing in-depth "how to" guidance on conducting studies or evaluations; summarizing the vast literature on research methodology and evaluation is beyond the scope of this paper.

This paper is organized in two parts. First, I survey the landscape of social accountability interventions by asking four questions: (1) What interventions are most common? (2) What contextual factors commonly have affected interventions? (3) Were the interventions effective? (4) How was causation attributed? These questions will help clarify the relationship between social accountability interventions and desired outcomes. Second, I examine assessment options at three programming stages: design, implementation, and closeout. I look at the available data sources, collection mechanisms and feasibility, indicator considerations and use, attribution issues, and sector-specific outcomes.

METHODOLOGY

I collected evaluations of social accountability interventions through Google Scholar searches and discussions with evaluation experts. I reviewed 58 articles for the criteria in Box 1. I found that 23 of these articles were general reviews of existing literature, and six others were not focused on social accountability. In total, 29 articles were included in the analysis at this stage. I then reviewed the bibliographies of the 23 reviews for articles that met our inclusion criteria and identified 20 further articles for inclusion. Adding these 20 studies to the previously identified 29 studies yielded a total of 49 articles reviewed for this study. I then classified each study by sector, research question, inputs, assessment methodology, and outcomes.

> **Box 1: Study inclusion criteria**
> ♦ Identification of a research question
> ♦ Stated research method
> ♦ Social accountability focus
> ♦ Peer-reviewed content, except for working papers
> ♦ Stated findings (positive or negative)
> ♦ Specific rather than general reviews

I recognize that the studies included do not represent the entire universe of research on social accountability. Mansuri and Rao (2013), for example, included nearly 500 articles in their work on participation in local development. While the review is not comprehensive, I am nonetheless confident that I have identified a representative universe of studies[1] that met our inclusion criteria.

SOCIAL ACCOUNTABILITY INTERVENTIONS

In this section, I present an analysis of specific interventions employed, the contextual factors that affected implementation, the effectiveness of these interventions, and the tools and indicators used to assess impact. I have sought to identify the common threads to paint a complete picture of social accountability interventions documented in the literature.

What interventions are most common?

I found significant clustering of interventions studied in the research literature. The five specific categories of interventions identified were: (1) use of media or community meetings to improve government transparency (with or without participation or responsiveness interventions); (2) community-driven development (CDD)[2]; (3) participatory budgeting and monitoring; (4) government sanctions,

---

[1] The reference list at the end of this paper is a good resource for practitioners seeking to learn more about evaluating social accountability programs.

[2] According to Dongier et al. (2003, p. 303): "Community-driven development (CDD) gives control of decisions and resources to community groups. These groups often work in partnership with demand-responsive support organizations and service providers including elected local governments, the private sector, NGOs, and central government agencies. CDD is a way to provide social and infrastructure services, to organize economic activity and resource management, to empower poor people, improve governance, and enhance security of the poorest."

audits, or monitoring (with or without community input); and (5) more democratic selection of leaders through elections.

Transparency interventions

Programs to strengthen transparency were very common interventions, often taking the form of publishing budget or performance information about government services in newspapers or other media outlets. These interventions isolated the effects of transparency by focusing solely on information sharing efforts—i.e., the implementers did not combine information sharing with community participation or with government oversight mechanisms, sanctions, or incentives. For example, one study sought to reduce corruption by publishing data on national government transfers to schools in newspapers, while another disseminated information about the performance of Members of Parliament (MPs), in order to provide information for elections (Humphreys & Weinstein, 2012; Reinikka & Svensson, 2004). Fewer studies combined transparency with incentives for government responsiveness or improved citizen participation. Bjorkman, de Walque, and Svensson (2014) and Hubbard (2007) both incorporated these elements into their studies of health facility committee effectiveness and of reductions in corruption around school grants via publishing financial information and incorporating a monitoring system for the grants. The most common differentiator among the transparency studies was the inclusion or exclusion of participation or responsiveness interventions.

Community-driven development

Researchers used a wide variety of mechanisms to study citizen engagement in CDD projects. Fearon, Humphreys, and Weinstein (2011) looked at the impact of elected councils on oversight of infrastructure. The program created these councils to manage specific CDD projects and mandated the participation of women. Casey, Glennerster, and Miguel (2013) reported on a scheme that used mandates to ensure gender equity on village development committees. These committees oversaw block grants for public goods, such as schools, latrines, grain storage, or skills training. Not all CDD projects supported infrastructure, however. Humphreys, de la Sierra, and van der Windt (2012) assessed the role of village committee management on unconditional cash transfer programs. Platteau and Gaspart (2003) recognized the ability of local elites to capture CDD mechanisms. They assessed the effectiveness of measures to discipline local leaders when they captured resources meant for the poor.

Participatory budgeting, planning, and monitoring

While CDD mechanisms seek to empower and engage citizens outside of normal government structures, governments can also engage citizens in budgeting, planning, and monitoring of services. Touchton and Wampler (2013) assessed the effects of Brazil's 20 years of experience with participatory budgeting, in which citizens present requests and identify investment priorities in a series of meetings, on spending on health and sanitation. Gonçalves (2014) similarly researched how Brazil's participatory budgeting mechanisms had affected infant mortality. The author noted that citizens could also participate in deciding which projects to implement and on what timeline. In Indonesia, researchers assessed the use of direct elections to bypass elite capture, hypothesizing that direct democracy would improve the political legitimacy of decision making (Olken, 2010). Citizen monitoring of government programs also was an important intervention in this compilation of research. For example, Barr and colleagues (2012)

studied an intervention to better coordinate monitoring among school management committees in Uganda. Finally, citizen engagement can also involve including previously marginalized groups. Holland et al. (2012) explicitly addressed engaging these groups in citizen forums to access entitlements and challenge social norms.

Sanctions, audits, and leadership selection mechanisms

The last two types of intervention groups studied the short and long routes of accountability (World Bank, 2003). The short route involves strengthening accountability through sanctions, audits, or monitoring of government officials, with varying levels of citizen engagement; the long route considers improving leadership selection through democratic and inclusive mechanisms. Among the interventions evaluating the usefulness of government sanctions, studies included three types of reforms: (1) improving government sanctioning systems without offering opportunities for community participation, (2) improving government sanctioning systems and including strong mechanisms for community participation, and (3) improving community participation without introducing any structural interventions to improve government sanctioning systems. Examples included government audits of programs and budgets, as well as CDD projects (Olken, 2005; Plateau & Gaspart, 2003). These studies looked primarily at technical improvements to government systems to hold officials accountable for effective and efficient programs, even if opportunities for citizen input were constrained. Of the second type of reform, studies assessed community monitoring of schools and teachers, in situations in which incentives were instituted for teacher performance and collective action to clean up trash, paired with sanctions for nonperformance that were imposed by government or traditional leaders (Duflo, Dupas, & Kremer, 2015; Sheely, 2013). Studies also reported on the effects of participation through social audits and citizen report cards, neither of which was accompanied by improvements to government mechanisms to sanction or monitor officials (Ravindra, 2004; Shankar, 2010).

Finally, creating more democratic and inclusive leadership selection processes was also a major theme. Chattopadhyay and Duflo (2004) assessed the influence of reserving spaces in local government for women, while Grossman (2014), Beath, Christia, and Enikolopov (2013a), and Baldwin and Mvukiyehe (2011) all measured the effects of election vs. selection of leaders on how public goods or project resources were allocated.

What contextual factors commonly affected interventions?

Regardless of the type of intervention studied, context matters. O'Meally (2013) noted that similar social accountability approaches can vary greatly among communities and across time, and that implementers must consider how to get beyond "best practice" to "best fit." Program success often depends on local factors, such as history, community trust, social capital, geography, and political systems (Booth, 2011; Mansuri & Rao, 2013). Contextual factors to consider can be found in Box 2. Joshi (2014) further examined the macro- and micro-levels of context. She looked at how accountability was embedded in both national-level politics and citizen–state relations, as well as local factors that influenced specific interventions—such as local elite capture, personalities of key personnel, or experience with previous social accountability mechanisms—even within broadly similar macro contexts. Contextual factors can also be important for researchers as they study specific interventions, especially if they are designing studies that rely on natural variation.

> **Box 2: Contextual factors for social accountability**
>
> **Macro factors**   **Micro factors**
>
> ♦ History   ♦ Institutional culture
> ♦ Gender norms   ♦ Personalities of key personnel
> ♦ Community trust   ♦ Experience with social
> ♦ Social capital   accountability mechanisms
> ♦ Citizen-state engagement   ♦ Elite capture
> ♦ Geography   ♦ Leadership selection
> ♦ Political systems   ♦ Availability of media
>
> Adapted from Joshi (2014); Booth (2011); and Mansuri and Rao (2013).

Local (or micro) factors

The research literature showed that a wide range of local (or micro) factors can impact program success. In Pakistan, Khwaja (2009) showed that preexisting community trust was a critical factor in maintaining commonly owned infrastructure. Communities feared that resources, financial or otherwise, would be misappropriated, and were therefore reluctant to support maintenance. He also found that existing community inequality affected maintenance. Increasing inequality by 0.1 units (on the Gini index of national income equality) lowered maintenance by 24 percentage points. Social capital is also a key factor for the success of social accountability, as evidence from Sri Lanka showed that farmer organizations with higher measures of social capital were able to coordinate inputs from government to improve harvest yields (Uphoff & Wijayaratna, 2000).

Preexisting processes for selecting leaders also can have a significant effect on social accountability outcomes. Baldwin and Mvukiyehe (2011) found that introducing participatory processes to closed systems for selecting leaders in Liberia led to greater collaboration between community members and leaders, improved civic engagement of citizens, and better alignment of citizen and leader interests. Caseley (2003) showed that reforms to complaint mechanisms relied heavily on the systems, structures, and leaders that already existed within an institution. In this case, the Hyderabad Metropolitan Water Supply and Sewerage Board was resistant to change, so the researchers tried to improve accountability by building on a preexisting complaints system. More broadly, Brett (2003, p. 1) claimed that demand-side measures can succeed only in areas where government leaders are already responsive, and that such interventions fail where "local conditions make cooperative and collective action very difficult." As later evidence upholding this point, Banerjee and colleagues (2010) studied improving community information and participation in Indian education, where according to a baseline study, community members were not aware of key mechanisms to hold school leaders accountable for educating children, nor had they complained about education services to government. In this study, they found that the interventions showed no improvement in community control over public schools.

Joint macro and micro factors

Among the studies reviewed, a few considered both macro and micro contextual issues. Fearon and colleagues (2011, p. 3), analyzing the use of community councils to guide reconstruction efforts in

postwar Liberia, recognized the macro social upheaval that had taken place during the war. In analyzing the micro context, however, he found "high levels of social capital and behavioral evidence of strong capacity for organizing and providing public goods" in the sample communities. Sheely (2013) also took into account the macro and micro contexts at play in rural Kenya. In studying how government and communities were working together to ensure trash removal, he found that preexisting conflict between local communities and government reduced the ability of the government and communities to maintain trash-free neighborhoods, both pre- and post-intervention.

Contextual variations

Finally, quasi-experimental designs commonly take advantage of contextual variations to study social accountability issues. For example, Beath, Christia, and Enikolopov (2013b) studied the effects of locally elected committees on food aid distribution in Afghanistan. Due to social norms restricting women's leadership, they were able to study the effects of mandating women's participation in village governance structures. They found increases in embezzlement and no changes in targeting or participation. Reinikka and Svensson (2004) used local variations in the distribution of newspapers to study the effects of the transparency of budgets and financial transfers on corruption, and found that improving transparency significantly reduced leakage of these resources.

Were the interventions effective?

Before identifying if programs were effective, it is necessary to define *effectiveness* in the context of social accountability. As noted previously, Brinkerhoff and Wetterberg (2015) identified three areas that social accountability interventions attempt to influence: governance and democracy, citizen empowerment, and service delivery. This section focuses on the types of social accountability interventions that improve outcomes across these three areas.

As we have seen in the previous sections, intervention effectiveness is driven by how programs are implemented and by the various factors at play in a given context. Fox (2015) highlighted three general takeaways in his literature review on the effectiveness of social accountability programs: information is not enough (i.e., disseminating data, without building a social movement or changing the incentives of politicians), bottom-up monitoring often lacks bite, and CDD projects are often captured by local elites. Recognizing these three analytical insights, I compare and contrast interventions that address each of these issues (the role of transparency, bottom-up vs. top-down accountability, and the impact of CDD) in turn. I also analyze the impact of studies that took advantage of multiple accountability pathways, including linking citizens and direct-line service providers, elected officials, and oversight agencies.

The literature shows a distinct difference between **"transparency-only" and "transparency-plus"** interventions, with evidence suggesting that effectiveness is enhanced when information-sharing is complemented by other social accountability activities. In Uganda, two studies, using the same data, came to different conclusions on whether publishing national grants to primary schools in newspapers reduced corruption. On one hand, Reinikka and Svensson (2004, p. 3) claimed that "improved access to information significantly reduced local capture." On the other hand, when Hubbard (2007, p. 9) analyzed the data, he found that "only in communities that were literate and assertive enough to act when abuse

was revealed were information disclosure measures effective." In a related report, the Royal Netherlands Embassy (2003) found that teachers often continued to steal grant funds following the transparency campaigns in Uganda. Overall, information disclosure did help to reduce leakage, but only in the context of increased government oversight, changes in funding systems, and new incentives for parental monitoring. Other research, including Olken's (2005) study of audits in Indonesia and Holland and colleagues' (2012) study of inclusion in Nepal, supports the conclusion that sharing information to prompt citizen voice is important, but that government reforms are also critical to ensuring social accountability: what Fox (2015, p. 352) called "voice plus teeth."

Bottom-up accountability mechanisms can be effective at improving sector-specific outcomes, but can falter if leaders do not have incentives to respond to citizen monitoring. In a study in three Indian states, Pandey, Goyal, and Sundararaman (2008) used a randomized controlled trial (RCT) to analyze the impact of improving information on school oversight, management, and the benefits of school. They found that this approach showed a positive impact on reading and math achievement, although they also found mixed results on teacher engagement and attendance. In the health sector, community monitoring of public health services was found to improve health outcomes, reducing under-five mortality by 33 percent (Björkman & Svensson, 2009). Holland and colleagues (2012, p. 195) took a broader approach: analyzing both supply- and demand-side accountability. Using a case study approach, they found that both were necessary to empowering historically excluded groups in Nepal, noting that "interventions have a more sustained impact when they tackle the institutions—national and sub-national governance systems and deep social structures—that underpin accountability relations." For citizen voice and monitoring to be effective, however, a responsive government is required. In a study of leadership selection mechanisms in Uganda, Grossman (2014) noted that directly elected leaders were more responsive to citizens than appointed leaders, lending support to the importance of long-route mechanisms. He also found that monitoring mechanisms were significantly less likely to be used if leaders were appointed, rather than elected. Olken (2005) identified a similar mechanism in road projects in Indonesia. He found that top-down expenditure audits reduced leakage by 8 percent, but that grassroots participation in monitoring reduced leakage only from wages, not from materials, which made up 75 percent of total expenditures. These findings echo Fox's (2015) conclusion that bottom-up monitoring must be paired with top-down systems to be effective.

Finally, I look at a key mechanism for holding service providers accountable: **CDD approaches**. These use community-run structures that fall outside of normal government mechanisms to identify community needs and allocate resources, and can have both positive and negative attributes. CDD, as a mechanism, is used to achieve widely different goals. Berman (2014) described how CDD was used to manage a conditional cash transfer (CCT) program. When a CDD approach was used to manage CCTs, families were 5 percent less likely to fall ill, while children under four were 10 percent less likely to get sick. The intervention had educational benefits as well: Girls were 24 percent more likely to complete primary school. In addition to service outcomes, the study showed social cohesion outcomes. Members of the treatment group were more likely to attend village meetings and to devote their time to community projects.

CDD also has been used to prioritize infrastructure investments, with mixed results. The GoBifo program in Sierra Leone provided block grants for community infrastructure, which were managed

community committees (Casey et al., 2013). The evaluators found improved program implementation, but no significant effect on social change, such as the way decisions were made, community trust, or participation in government. The Tuungane program in the Democratic Republic of Congo also used block grants to build infrastructure, but its evaluators found little evidence of impact. Treatment and control groups showed similar levels of transparency, fund misappropriation, and measures of trust, although they did find some evidence that citizens in treatment groups were more likely to complain if funds were misused (Humphreys et al., 2012).

There has been considerable debate over the extent of elite capture in CDD programs. For example, Fox's (2015, p. 348) review of social accountability found that "community driven development programs are often captured by local elites." Esman and Uphoff (1984, p. 249) had reached the same conclusion 20 years earlier, noting that:

> The most prominent members are invariably selected and then given training and control over resources for the community.... The more training and resources they are given, the more distance is created between leaders and members. The shortcut of trying to mobilize rural people from outside through leaders…is likely to be unproductive or even counterproductive, entrenching a privileged minority and discrediting the idea of group action for self-improvement.

Platteau and Gaspart (2003), however, found this claim to be a broad generalization, noting that community empowerment and committee selection methods can make CDD more or less vulnerable to elite capture. They also proposed a leader-disciplining mechanism to overcome elite capture. This mechanism would use fraud detection and disburse funding in tranches to hold leaders accountable. Platteau and Gaspart (2003) also highlighted a major implementation challenge inherent with CDD: Donors often rush into CDD programs without concurrently building up democratic institutions, leading to significant elite capture. They argued that contextual issues, such as weak local community trust or leadership, drive elite capture.

A final key point is that **successful interventions take advantage of multiple accountability relationships:** between citizens and direct-line service providers, citizens and elected officials, and also oversight agencies and providers. These relationships often cross the supply/demand divide, allowing interventions to take advantage of different types of entry points. Caseley (2003, p. 3), in a study of water provision in Hyderabad, India, noted that "multiple accountability relationships, operating between external actors and Metro Water staff, have collectively contributed to sustained organizational change and improved service delivery performance." He found that three months after Metro Water developed a systematic complaint mechanism, it had increased the number of monthly new connections by 79 percent, even though no new staff or resources had been devoted to the utility. He explained these improvements as the result of both demand-side (active citizen engagement) and supply-side (formal complaint systems) accountability mechanisms.

How is causation attributed?

Identifying causation, or even plausible attribution, is a critical concern for social accountability research. Of the literature reviewed, the majority of analyses applied quasi-experimental and randomized

designs, which looked for statistically significant differences between treatment and control groups. Cross-sectional studies were also common, although these studies were employed mostly to analyze correlation, rather than causation. While these approaches are useful for highlighting *what* works, they are often unable to explain *why* it works. For example, one quasi-experimental study showed that audits reduced corruption in Indonesia, but the causal pathways could only be hypothesized (Olken, 2005).

Further investigation, using qualitative designs, can illuminate causal pathways and map processes (Barnes & Weller, 2012). Qualitative approaches that analyze research questions from a variety of viewpoints can provide these causal pathways. Ravindra (2004), for example, used cross-sectional surveys to assess citizen satisfaction with government services, and then interviewed key informants to better understand how civil society was using those surveys to lobby government.

Of course, development research, generally, and social accountability research, specifically, tread a fine line on causality, especially when trying to link accountability mechanisms to service-delivery outcomes. Baur et al. (2001, p. 2) argued that "significant factors make it impossible to isolate the effects of a single development intervention," and that an "attribution gap" exists between direct intervention effects and "highly aggregated progress," which is influenced by other, uncontrollable, factors. They claimed that "plausible linkages between the project, program or strategy under investigation, and the impact observed on the ground" should be considered the gold standard for evaluation. They also identified seven standards for identifying these plausible linkages (see Box 3).

Woolcock (2009, p. 1) went further in calling for a broader understanding of causation in development, claiming that "the development community…is prone to routinely making attribution errors" as a result of a lack of knowledge of how inputs lead to outcomes. He noted that social accountability programs, such as community-driven approaches, violate many assumptions of experimental designs, such as that treatment and control groups are similar, or that the treatment is uniform, instead of continually adapted to context by front-line workers. Scaling up or transferring interventions to new settings is also problematic. For example, in moves from "small empowerment projects" to nationally relevant programs, political-economy factors become more prominent, changing the interventions substantially (Woolcock, 2009, p. 6). Contextual factors such as community trust and state legitimacy may affect whether social

> **Box 3: Elements for establishing the plausibility of impact**
>
> 1. Source of the impact being investigated
> 2. Model of impact used by the impact evaluators and how it applies to the case
> 3. Objectives of the impact assessment
> 4. Theory of action on which the intervention or strategy has been based
> 5. Hypotheses tested by the impact assessment
> 6. Other factors that could have affected the observed changes
> 7. Informed opinions that support and contest the study findings
>
> Source: Baur et al. (2001), p. 2.

accountability programs can be implemented at all. Stern et al. (2012, p. 8) also identified this problem, stating that "the contextual characteristics in which both control and treatment groups are embedded in [sic] might influence their response to the intervention, thus impairing the external validity."

EVALUATING[3] EFFECTS OF SOCIAL ACCOUNTABILITY INTERVENTIONS

Given this understanding of what social accountability research exists and what it shows, I now move to how social accountability interventions can be evaluated. This section suggests to program managers tools to evaluate interventions, while also recognizing that programs do not always take place as planned. Programs are often implemented without a baseline evaluation, control groups may not be tracked, activities change mid-program, and donor demands can change evaluation questions. As a result, evaluations that track change over time or precisely compare intervention and control groups may not be possible. Acknowledging these constraints exist, I identify key considerations for conducting social accountability research at the design, implementation, and evaluation stages of the program cycle, across five factors: **outcome types, possible indicators, strength of attribution, study design,** and **data collection.** These considerations assume that the program under discussion is already designed—i.e., that activities are set, a budget is estimated, and the geographical and technical scope of the program are clear.

> *"The best time to plant a tree was 20 years ago. The second best time is now."*
>
> *—Chinese proverb*

Design

Beginning program evaluations during the initial design is clearly the most desirable option, as this stage provides the most opportunities for developing baseline data, organizing control groups, and defining inputs and outcomes. For academic research on social accountability, where the main goal is the knowledge generated by the study, the scope of evaluation measures will be different from a donor-funded scale-up of a social accountability mechanism or tool. For program management purposes, we are often interested in a more narrow set of questions than an academic study would be. This section therefore focuses on program evaluations, rather than on more rigorous academic studies.

Types of outcomes

The first factor to consider when evaluating a social accountability program is the **type of outcome** to measure. Deciding which outcomes are important to the intervention depends heavily on the intervention. For example, complaint surveys, service charters, and citizen review boards often focus on improving government services in specific sectors. As a result, appropriate measurements could include changes in school attendance or immunization rates from a baseline measure. Relevant governance and empowerment outcomes—as a result of civic education, community mobilization, or transparency campaigns—could include public investment changes, participation of marginalized populations in decision making, or electoral outcomes.

Key questions for deciding outcomes of interest:

♦ What are the program aims?
♦ What is the indicator measuring? (i.e., service improvements? participation?)

---

[3] For this paper, I have used the terms "evaluate" and "assess" interchangeably.

♦ Is the outcome a realistic effect of the program?

♦ Is there a legitimate causal pathway from program to outcome?

Possible indicators

With available time and money, the universe of **possible indicators**[4] to measure outcomes is quite large. In order to narrow the choice of indicators, Gajda and Jewiss (2004, para. 6) recommended thinking about indicators for program evaluation as "the gauge of whether, and to what degree, you are making progress." The choice of indicators will also be driven by a number of technical and programmatic factors. For example, will the program need to count inputs (e.g., number of people trained) or outputs (e.g., number of people served)? Managers also need to consider the appropriate mix of quality indicators, which measure how the interventions was achieved; and quantity indicators, which measure the number of people affected by the intervention. The availability of data, either through routine mechanisms or one-off surveys, can also drive indicator selection. Context-specific issues, such as distance to service sites, openness to surveys, willingness to be interviewed, and availability of sampling data may also make collection of certain indicators challenging.

Key questions for indicator development:

♦ Is the indicator measurable given program constraints (e.g., money, distance, culture)?

♦ Are resources available to conduct a baseline? Or to track a control group?

♦ What is the appropriate mix of quality and quantity indicators?

♦ Which inputs should we measure? Which outputs?

♦ Do we have access to needed government or survey data?

Attribution of causation

The third factor to consider at the design stage is **feasibility of attributing causation.** As with many other factors, creating an evaluation process that can determine causality is much easier at the design stage, due to the ability to measure differences over time. If we are evaluating an activity to justify public spending or to hold governments accountable for results, then simply showing that an input led to a desired outcome, without a clearly defined theory of change, might suffice. This type of evidence can be obtained through a quantitative study. However, evaluators often want to know why an intervention worked, not just that it worked. Qualitative information is often required, via in-depth interviews or focus groups, to generate a theory of change that explains the causal pathway for the intervention. Glennan (1996, p. 64) claimed that the theory of change is critical to understanding, arguing that "two events are causally connected when and only when there is a mechanism connecting them" and "the necessity that distinguishes connections from accidental conjunctions is to be understood as deriving from an underlying mechanism."

We must also consider what types of threats to establishing causation we are willing to accept. Even gold-standard RCTs are subject to threats to validity. They cannot control for unmeasurable

---

[4] Annex A contains a table of illustrative indicators for use at the design stage.

contextual differences between treatment and control groups, influence between treatment and control groups, or homogeneity of the intervention across all treatment groups. Additionally, the intervention may not work in the same way throughout the life of the program, without being changed midstream to suit donor, community, or technical needs (Woolcock, 2009). Other methods, such as qualitative measures, must be used to address these concerns. Finally, various constraints can affect the ability to identify causation. Availability of time, funding, appropriate interviewees, control groups, or political support can all affect efforts to build the "perfect" evaluation.

Key questions for understanding the strength of attribution:

♦ Why do we need to understand causation?
♦ Is a theory of change or critical pathway analysis required?
♦ Do we need to understand how to improve the intervention? Or what supporting/contextual factors are necessary for the intervention?
♦ What threats to validity is the evaluator willing to accept? How can they be mitigated?

Study design

Using the information identified in the first three factors, the fourth consideration is determining **the appropriate study design** (see Table 1). If the evaluation process begins at the design stage, it is technically feasible to collect baseline data or identify a control group. The decision to do so, however, requires that a donor be willing to wait for collection of such data; it also hinges on whether a treatment can be randomized, what costs are involved, and what types of bias we are willing to accept. An RCT or quasi-experimental design can be a powerful tool for understanding whether an intervention is effective, as randomization controls for selection bias. When conducting an RCT or quasi-experimental study is not feasible because of cost or other constraints, a cross-sectional study for baseline data before the intervention is conducted can be a cost-effective option. Interviewing this same group of people post-intervention can yield powerful longitudinal data, even without a control group.

Key questions for study design selection:

♦ Is it possible to find and track the same group of people at different times? (longitudinal)
♦ Can the intervention be randomized? If not, is a control group feasible? Will the context allow randomization? (RCT/quasi-experimental study)
♦ Are resources or time available to do a longitudinal or quasi-experimental study? If not, is a qualitative study or panel study more appropriate?
♦ What ethical considerations[5] are involved?

---

[5] The Belmont report (National Commission for the Protection of Human Subjects, 1978) identified three principles. (1) Respect for Persons: Research subjects must have the ability to consider alternatives, decide on their level of participation, and act without influence. (2) Beneficence: This means both not harming research subjects and doing whatever is possible to protect research subjects from harm. (3) Justice: The costs and benefits of the research should be fairly distributed.

♦ How quickly does the funding agency want to begin implementation? (i.e., how long is it willing to wait to collect baseline information)

♦ What biases (e.g., selection or reporting bias) are most likely?

Table 1: Matrix of potential study designs for social accountability

Design	Timing	Relative funding needs	Potential tools	Estab-lishes causality?	Other major advantages	Other major disadvan-tages
Cross-sectional study[a,c]	Closeout	$$	Survey, knowledge tests	No	Is quick, represen-tative	Has inherent self-report bias
Qualitative study (including case studies)[a]	Design, implemen-tation, and/or closeout	$	In-depth interviews, focus groups, community mapping, social network analyses, scorecards, direct observations, pile sorting	Depends	Provides depth and detail	Produces results that are difficult to generalize
Randomized controlled trial[b]	Design	$$$	Survey, routine data	Yes	Controls bias	Produces results that are difficult to generalize
Quasi-experimental study[b]	Design	$$$	Survey, routine data	Yes	Uses "natural experiments"	Has inherent allocation bias
Longitudinal (cohort or panel) studies[c]	Design, implemen-tation	$$$/$$	Survey, routine data	Yes	Is useful to find inci-dence, represen-tative	Is time-consuming, risks attrition, and conditions respondents to questions
Case-control study[c]	Closeout	$	Existing case studies	No	Can study dynamic processes	Has inherent recall bias
Mixed methods[b]	Design, implemen-tation, or closeout	$$	Survey, any qualitative tool	Yes	Offers greater breadth of understan-ding; explains causal processes	Generates research designs and analyses that can be complex

[a] Descriptive; [b] Experimental; [c] Observational.

Source: Adapted from Aslam et al., 2012; Grimes and Schulz, 2002; Lewallen and Courtright, 1998; Maxwell, 2004a, 2004b; and Plümper, Troeger, and Neumayer, 2010.

The final consideration is how to **collect the information necessary** to measure the chosen indicators. Once a study design is chosen, identifying data collection methods at the design stage is fairly straightforward. Almost all designs will require a baseline study of some type, regardless of the existence

of a control group. For surveys, a unit of analysis (e.g., household, person, organization) should be determined. The sample size of the baseline, as well as methods for randomization, can be calculated using a number of online tools.[6] Routine data can also serve as a source of baseline data: Demographic and Health Surveys (see USAID's DHS information page, http://www.dhsprogram.com), utilization information, and electoral information can all provide valuable baseline data for social accountability activities. For both qualitative and quantitative methods, the same tools should be used at baseline as in the final study.

Key questions for data collection:

♦ What data are already available?
♦ What tools are appropriate for the study design chosen?
♦ What should be the sample size of a survey?
♦ If randomization is necessary, how will treatment and control groups be assigned?

Implementation

If a program is already in the middle of implementation and no evaluation method has been determined, a number of options for measurement have already been closed off. Since the interventions have begun, any study requiring a baseline—i.e., an experimental design, a panel, or cohort study—is not possible. Multiple options are, however, still available for measuring the intervention process or change until the end of the program.

As with the design stage, determining the outcomes of interest and the indicators to measure those outcomes are critical. While the outcomes of interest are likely similar at both the design and implementation stages, the indicators used to measure those outcomes must change. For example, it is no longer possible to measure service utilization changes from a set baseline. However, if routine data systems have captured an outcome of interest—such as test scores—at

> **Evaluation as programming: A Ugandan case study**
>
> In Uganda, the Governance, Accountability, and Participation Program (GAPP) is strengthening local government accountability for improved service delivery. The program team works with local civil society organizations to improve communication with local government and to analyze government actions. Before evaluating the effectiveness of the approach, GAPP recognized a number of constraints. For example, GAPP does not have control districts; and standardizing interventions across diverse sites is not appropriate. As a result, project complexity and interconnections among interventions make attribution difficult.
>
> To measure implementation effect, GAPP uses an iterative approach. It begins by pilot testing interventions in a few districts to adapt inputs before scale-up. Pilot tests are measured through in-depth case studies and learning events. They use indicators that focus on improving the enabling environment, such as measures of engagement with both citizens and the private sector, strengthened local government planning and financial management, and improved leadership skills among local government officials. Qualitative indicators are measured through "outcome mapping journals" used by field workers to document lessons learned during implementation.

regular intervals, it may be possible to use a previous data collection point as a baseline and measure

---

[6] For free online tools to assist with sample size calculation, see DSS Research (2015); for randomization, GraphPad Software, Inc. (2015).

changes in future years as indicative of a program effect. As a result, routinely collected service delivery indicators may be more feasible than customized governance and empowerment indicators.[7]

On the other hand, it is still possible to collect input or process indicators that measure program implementation, rather than outcomes. Even though some inputs may have been missed between design and the start of the evaluation process, these data can still measure progress toward social accountability outcomes. Quantitative input indicators, such as counting the number of citizen complaints, people trained to track budgets, or participants in school meetings, can be usefully linked to output indicators, such as test scores, corruption perceptions, or service utilization, but will not capture the full scope of program implementation. These indicators can be useful in reconstructing a longitudinal study, such as cohort[8] or panel design, depending on the program participants. Unfortunately, determining causality may not be possible under these circumstances, as uncontrolled factors may affect the indicators of interest, or the relevant change may have already taken place.

If no control group was (or can be) established, qualitative studies are often the best study designs for understanding causation. Concept mapping or pathway analysis can lead to a case study with an embedded theory of change.[9] Often, in-depth interviews or focus group discussions are used to understand the strength of the causal relationships identified in the concept maps or theory of change. These methods, while not revealing statistical significance, can deliver explanations for how a social accountability input leads to a specific output. Both internal and external validity challenges are associated with these methods. The observer-expectancy effect (i.e., researcher cognitive bias unconsciously influences the participants) can change the responses of interviewees,

> **Developing causal chains in Indonesia**
>
> In Indonesia, the Expanding Maternal and Neonatal Survival (EMAS) program is evaluating the effect of "local government and civil society accountability, support, and advocacy on maternal and neonatal health outcomes." Although the program started in 2010, EMAS used cross-sectional surveys and causal pathway analysis to understand the effects of the first phase of the program.
>
> The cross-sectional surveys involved interviews with facility staff, health providers, and local government leadership. A midpoint survey in 2015 also was conducted and will be supplemented by an endpoint survey in 2016 to understand provider knowledge, clinical governance, and referral practices. Additionally, in-depth interviews and focus group discussions will be used to conduct a causal pathway analysis to explain "how the program's major interventions contribute to program outcomes." Furthermore, the causal pathway analysis will articulate the theory of change developed at program start-up and will "highlight…evidence that change is occurring." The pathways could also be used to inform scale-up in subsequent phases of the program.
>
> The evaluation was not complete as of this writing (mid-2015) and findings were not yet available.

confirmation bias can affect how researchers interpret responses, recall bias can affect the quality of data, and potential confounding variables may not be apparent and must be controlled (Cook, 2010; Plous,

---

[7] Annex B contains a table of illustrative indicators for use at the implementation stage.

[8] Cohort designs may be either prospective or retrospective, depending on whether outcomes of interest have already been achieved.

[9] For further reading, Stein and Valters (2012) considered the application of theories of change to international development programs, while Daley (2004) assessed how concept maps are used in qualitative research. Pathway analysis is described in both Barnes and Weller (2012) and Weller and Barnes (2014).

1993; Weller & Barnes, 2014). Finally, Myers (2000) affirmed the view that qualitative research often is not applicable outside of the specific context in which the study was done.

Considering the limited options for studying social accountability once program implementation has begun, data collection mechanisms are similarly constrained. For longitudinal studies, obtaining data from preexisting data sets may be a possibility. As with the design stage, Demographic and Health Surveys, utilization information, and electoral information may be good sources of these data. For qualitative studies, in-depth interviews are best conducted with interviewees with strong knowledge of the process used to implement the program. Both internal and external viewpoints are critical. Focus group discussions are often best conducted with program beneficiaries. Boyce and Neale (2006) provided a strong overview of the use of in-depth interviews for development research, while Debus (1988) highlighted similar considerations for focus group discussions.

Closeout

When a development program has a year or less of implementation, managers often scramble to show that their program has had a positive impact. This dynamic is no different for social accountability activities. While drawing strong causal relationships at this stage is difficult, a few options do remain.

Unlike at the design and implementation stages, selecting predetermined outcomes of interest to measure change is not as critical to evaluating program success. This is not to say that outcomes and indicators aren't important. Instead, the focus of the evaluation efforts should be on identifying what outcomes *actually* occurred,

---

Kinerja: Retrofitting research questions

The USAID-funded Kinerja program was awarded in 2010 to improve local government service delivery. In 2013, USAID asked the program to conduct an evaluation to better understand (1) the program's results chains and (2) the likelihood of impact on school-based management, even though neither research question had been planned in advance.

The evaluation used qualitative methods to answer the research questions. Indicators were limited to quantitative ones, e.g., "the degree to which stakeholders consider Kinerja support to have facilitated achievements that could lead to future impact." Due to the timing of the data collection efforts, the evaluation focused on schools that had already showed success, to identify the likelihood of impact, creating a pro-Kinerja bias.

The evaluators found that school managers had learned how to develop plans and budgets, but they were unable to determine the sustainability of those skills. In contrast, the evaluators were able to comment on how sustainability could be achieved, such as introducing school-based mentoring and teaching district staff management and oversight skills. Finally, they found that managers emphasized infrastructure over human resource needs, in contrast to the management-based approach of the program.

Source: Social Impact, Solidaritas, and RTI International (2013).

---

rather than trying to identify what *should have* occurred. Woolcock (2009, p. 7) noted that "seasoned project managers should have a good sense of how long and in what form the impacts associated with a particular project in a particular context should take to materialize." Potential indicators for identifying program impact thus should draw on the experience of the managers. What indicators are likely to have been affected by the activity? What is possible to measure? Measuring quantity indicators at this stage is likely not possible, and the evaluation should focus on qualitative indicators, such as social norms challenged, maintenance quality, and community satisfaction.[10]

---

[10] Annex C contains a table of illustrative indicators for use at the closeout stage.

As with the implementation stage, collecting routine data may still be an option, if relevant information is available, for a panel study. Additionally, a cross-sectional study to draw correlations between inputs and outcomes is possible, while retrospective qualitative designs, such as case, or case-control, studies may provide insights into causal links. A retrospective cohort study is also technically possible, but requires a great deal of information about specific individuals and is unlikely to be feasible. Case-control studies can reconstruct inputs and outcomes, if an appropriate control group can be described. Without a control group, a case study, using in-depth interviews and/or focus group discussions, can still describe the social accountability inputs, the process used to implement the program, and the outcomes achieved. The case study model does not produce the same level of evidence as cohort studies or experimental designs, but can still allow researchers to draw causal links through qualitative data analysis. Additionally, case studies, and case-control studies, are subject to the same biases at this stage as in the implementation stage.

CONCLUSION

This study has examined existing research to pinpoint key concerns for linking social accountability inputs to various types of outcomes. In reviewing this research, I found the following:

1. **Attributing single vs. multiple interventions.** Many interventions attempted to isolate the effects of one social accountability intervention, whether it was a transparency campaign, a set of government sanctions, community-driven development, or participatory governance. Unfortunately, these studies often did not find strong program effects, or they were contradicted by subsequent studies. The studies that did conduct multiple interventions at the same time were often more effective, but attributing outcomes to specific inputs was difficult in programs with multiple inputs. In such instances, qualitative data are needed to explain the research findings.

2. **Incorporating contextual factors.** Researchers considered contextual factors to explain variations in program impact and often used these factors to design quasi-experimental studies. RCTs often attempted to control for some aspects of local context, such as wealth, population, and leadership. Few studies, however, conducted the deep dive into local context that O'Meally (2013) recommended. Many studies were subject to external validity threats, as a result of not being able to control for local context, if such a thing is even possible.

3. **Learning more about causation and social accountability.** Mapping causation for social accountability inputs is a difficult proposition. Researcher reliance on RCTs and quasi-experimental designs has led program managers to have more knowledge about what works in specific contexts, but we are often still in the dark about why interventions work, or the generalizability of findings. Without better qualitative information on how context affects outcomes, or even well-defined theories of change for how program inputs affect outcomes, causal pathways are likely to remain ill-illuminated.

The paper also has highlighted operational considerations for evaluating social accountability activities at various stages of the program cycle. I have noted that evaluation options can be significantly limited by a number of factors: timing, donor constraints, and the availability of routine data, among

them. Building on the findings from the first section of this paper, I have argued for the importance of using qualitative data to explain the findings of experimental or observational studies. I also have provided a range of questions for managers to consider as they evaluate social accountability programs, regardless of the stage in the program cycle. Potential indicators for evaluation at each stage of the program also have been suggested. A final consideration is how programs change over time. Managers are constantly tweaking interventions and approaches in response to donor, community, or technical needs. As a result, replicating the exact social accountability intervention, even with strong documentation, is impossible and causal pathways are likely to contain gaps in knowledge.

Ultimately, the demand for social accountability programs is only growing stronger. Donors and governments are seeking new ways to spur development, and finding that social accountability interventions, broadly, work. The race is on to find the mix of interventions, and contextual considerations, for improving service delivery, governance and democracy, and empowerment outcomes. As a result, the divide between research that discovers new interventions and programs that implement known "best practices" will only shrink, as donors require programs to show impact beyond counting beneficiaries. Managers must be ready to propose new evaluation approaches that can answer key questions posed by donors and governments, and respond to the needs of citizens everywhere.

REFERENCES

Aslam, S., Georgiev, H., Mehta, K., & Kumar, A. (2012). Matching research design to clinical research questions. *Indian Journal of Sexually Transmitted Diseases and AIDS, 33*(1), 49–53.

Baldwin, K., & Mvukiyehe, E. (2011). *The effects of participatory processes for selecting leaders: Evidence from changes to traditional institutions in Liberia.* New York, NY: Columbia University. Retrieved from http://www.columbia.edu/~enm2105/docs/liberia/clan_chief_selection_liberia.pdf

Banerjee, A.V., Banerji, R., Duflo, E., Glennerster, R., & Khemani, S. (2010). Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy, 2*(1), 1–30.

Barnes, J., & Weller, N. (2012). Finding pathways: Case selection for studying causal mechanisms in mixed-methods research. *SSRN Electronic Journal.* Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2150372

Barr, A., Mugisha, F., Serneels, P., & Zeitlin, A. (2012). *Information and collective action in community-based monitoring of schools: Field and lab experimental evidence from Uganda.* Working Paper. Hanover, NH: Dartmouth College. Retrieved from http://cega.berkeley.edu/assets/cega_events/3/doc7Zeitlin-et-al_Information-and-Collective-Action-in-Community-based-mMonitoring-of-Schools.pdf

Baur, H., Bosch, M., Krall, S., Kuby, T., Lobb-Rabe, A., Schutz, P., & Springer-Heinze, A. (2001). *Establishing plausibility in impact assessment.* Eschborn, Germany: GTZ. Retrieved from http://knowledge.cta.int/Dossiers/S-T-Policy/Measuring-the-impact-of-agricultural-research/Relevant-publications/Establishing-plausibility-in-impact-assessment

Beath, A., Christia, F., & Enikolopov, R. (2013a). *Do elected councils improve governance? Experimental evidence on local institutions in Afghanistan.* Working Paper. London: International Growth Centre. Retrieved from http://www.theigc.org/wp-content/uploads/2014/09/Beath-Et-Al-2013-Working-Paper.pdf

Beath, A., Christia, F., & Enikolopov, R. (2013b). Empowering women through development aid: Evidence from a field experiment in Afghanistan. *American Political Science Review*, *107*(03), 540–557.

Berman, D. (2014). *Tanzania: Can local communities successfully run cash transfer programs? From evidence to policy.* Washington, DC: Office of the Chief Economist, Human Development Network, World Bank. Retrieved from http://documents.worldbank.org/curated/en/2014/03/19278238/tanzania-can-local-communities-successfully-run-cash-transfer-programs

Björkman, M., & Svensson, J. (2009). Power to the people: Evidence from a randomized field experiment on community-based monitoring in Uganda. *Quarterly Journal of Economics, 124*(2), 735−769.

Bjorkman, M., de Walque, D., & Svensson, J. (2014). *Information is power: Experimental evidence on the long-run impact of community based monitoring.* Policy Research Working Paper 7015. Washington, DC: World Bank. Retrieved from http://www.povertyactionlab.org/publication/information-power-experimental-evidence-long-run-impact-community-based-monitoring

Booth, D. (2011). Towards a theory of local governance and public goods provision. *IDS Bulletin*, *42*(2), 11–21.

Boyce, C., & Neale, P. (2006). *Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input.* Watertown, MA: Pathfinder International. Retrieved from http://www2.pathfinder.org/site/DocServer/m_e_tool_series_indepth_interviews.pdf

Brett, E. A. (2003). Participation and accountability in development management. *The Journal of Development Studies*, *40*(2), 1–29.

Brinkerhoff, D. W., & Wetterberg, A. (2015). Gauging the effects of social accountability on services, governance, and citizen empowerment. *Public Administration Review*.

Caseley, J. (2003). *Blocked drains and open minds: Multiple accountability relationships and improved service delivery performance in an Indian city.* IDS Working Paper Series, 211. Brighton, UK: Institute for Development Studies. Retrieved from http://www2.ids.ac.uk/futurestate/pdfs/wp211.pdf

Casey, K., Glennerster, R., & Miguel, E. (2013). *The GoBifo project evaluation report: Assessing the impacts of community-driven development in Sierra Leone.* Prepared under Open Window 1 grant OW1.65NewDelhi: International Initiative for Impact Evaluation (3IE). http://www.povertyactionlab.org/publication/gobifo-project-evaluation-report-assessing-impacts-community-driven-development-sierra-l

Chattopadhyay, R., & Duflo, E. (2004). Women as policy makers: Evidence from a randomized policy experiment in India. *Econometrica*, *72*(5), 1409–1443.

Cook, C. (2010). Mode of administration bias. *Journal of Manual & Manipulative Therapy, 18*(2), 61–63.

Daley, B. J. (2004). Using concept maps in qualitative research. In A.J. Cañas, J. D. Novak, & F. M. Gonzales (Eds.), *Proceedings of the First International Conference on Concept Mapping, Pamplona, Spain, September 14–17, 2004* (pp. 191–199). Tallahassee: Institute for Human and Machine Cognition (IHMC), Florida State University.

Debus, M. (1988). *Methodological review: A handbook for excellence in focus group research.* Washington, DC: Academy for Educational Development.

Dongier, P., Van Domelen, J., Ostrom, E., Ryan, A., Wakeman, W., Bebbington, A., Alkire, S., Esmailo, T., & Polski, M. (2003). Community-driven development. Chapter 9 in *PRSP Sourcebook,* Volume 1. Washington, DC: World Bank. Retrieved from http://siteresources.worldbank.org/INTPRS1/Resources/383606-1205334112622/5805_chap9.pdf

DSS Research. (2015). Calculators: Researcher's toolkit [Online calculator]. Fort Worth, TX: DSS Research. Retrieved from https://www.dssresearch.com/KnowledgeCenter/toolkitcalculators/statisticalpowercalculators.aspx

Duflo, E., Dupas, P., & Kremer, M. (2015). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools. *Journal of Public Economics*, *123*, 92–110.

Esman, M. J., & Uphoff, N. T. (1984). *Local organizations: Intermediaries in rural development.* Ithaca, New York: Cornell University Press.

Fearon, J. D., Humphreys, M., & Weinstein, J. M. (2011). *Democratic institutions and collective action capacity: Results from a field experiment in post-conflict Liberia.* Paper presented at the 2001 Annual Meeting of the American Political Science Association (APSA), Seattle, Washington, September 1–4, 2011. Palo Alto, CA: Stanford University. Retrieved from http://www.columbia.edu/~mh2245/papers1/FHW_2012.pdf

Fox, J. A. (2015). Social accountability: What does the evidence really say? *World Development, 72,* 346–361.

Gajda, R., & Jewiss, J. (2004). Thinking about how to evaluate your program? These strategies will get you started. *Practical Assessment, Research and Evaluation*, *9*(8). Retrieved from http://PAREonline.net/getvn.asp?v=9&n=8

Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, *44*(1), 49–71.

Gonçalves, S. (2014). The effects of participatory budgeting on municipal expenditures and infant mortality in Brazil. *World Development*, *53*, 94–110.

GraphPad Software, Inc. (2015). QuickCalcs [Online calculator]. La Jolla, CA: GraphPad. Retrieved from http://graphpad.com/quickcalcs/randomize1.cfm

Grimes, D. A., & Schulz, K. F. (2002). An overview of clinical research: the lay of the land. *The Lancet*, *359*(9300), 57–61.

Grossman, G. (2014). Do selection rules affect leader responsiveness? Evidence from rural Uganda. *Quarterly Journal of Political Science,* 2014, 9, 1–44.

Health Finance and Governance. (2014). *Generating evidence of governance contributions to health outcomes. Workshop summary report, June 23, 2014.* Prepared for the USAID Global Health Bureau under the Health Finance and Governance (HFG) Project. Bethesda, MD: Abt Associates. Retrieved from http://pdf.usaid.gov/pdf_docs/PA00K4G3.pdf

Holland, J., Ruedin, L., Scott-Villiers, P., & Sheppard, H. (2012). Tackling the governance of socially inclusive service delivery. *Public Management Review*, *14*(2), 181–196.

Hubbard, P. (2007). Putting the power of transparency in context: Information's role in reducing corruption in Uganda's education sector. *SSRN Electronic Journal.* Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1100131

Humphreys, M., de la Sierra, R. S., & van der Windt, P. (2012). *Social and economic impacts of Tuungane: Final report on the effects of a community driven reconstruction program in Eastern Democratic Republic of Congo.* Impact Evaluation Report 7. New Delhi: International Initiative for Impact Evaluation (3IE). Retrieved from http://www.oecd.org/countries/democraticrepublicofthecongo/drc.pdf

Humphreys, M., & Weinstein, J. (2012). *Policing politicians: Citizen empowerment and political accountability in Uganda – Preliminary analysis.* Working paper. London: International Growth Centre, London School of Economic and Political Science. Retrieved from http://www.theigc.org/wp-content/uploads/2015/02/Humphreys-Weinstein-2012-Working-Paper.pdf

Joshi, A. (2014). Reading the local context: A causal chain approach to social accountability. *IDS Bulletin, 45*(5), 23–35. Brighton, UK: Institute of Development Studies.

Khwaja, A. I. (2009). Can good projects succeed in bad communities? *Journal of Public Economics*, *93*(7), 899–916.

Lewallen, S., & Courtright, P. (1998). Epidemiology in practice: Case-control studies. *Community Eye Health*, *11*(28), 57–58.

Mansuri, G., & Rao, V. (2013). *Localizing development: does participation work?* World Bank policy research report. Washington, DC: World Bank. Retrieved from http://dx.doi.org/10.1596/978-0-8213-8256-1

Maxwell, J. A. (2004a). Using qualitative methods for causal explanation. *Field methods*, *16*(3), 243–264.

Maxwell, J. A. (2004b). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, *33*(2), 3–11.

Myers, M. (2000). Qualitative research and the generalizability question: Standing firm with Proteus. *The Qualitative Report, 4*(3/4), 1–9.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. US Department of Health, Education, and Welfare (DHEW) Publication No. (OS) 78-0014. Washington, DC: DHEW. Retrieved from http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html

Olken, B. A. (2005). *Monitoring corruption: Evidence from a field experiment in Indonesia.* NBER Working Paper No. 11753. Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w11753

Olken, B. A. (2010). Direct democracy and local public goods: Evidence from a field experiment in Indonesia. *American Political Science Review*, *104*(02), 243–267.

O'Meally, S. C. (2013). *Mapping context for social accountability: A resource paper.* Washington, DC: Social Development Department, World Bank. Retrieved from http://siteresources.worldbank.org/EXTSOCIALDEVELOPMENT/Resources/244362-1193949504055/Context_and_SAcc_RESOURCE_PAPER.pdf

Pandey, P., Goyal, S., & Sundararaman, V. (2008). *Community participation in public schools: The impact of information campaigns in three Indian states.* World Bank Policy Research Working Paper Series. Washington, DC: World Bank. Retrieved from https://openknowledge.worldbank.org/handle/10986/6359

Platteau, J. P., & Gaspart, F. (2003). *The "elite capture" problem in participatory development.* Namur, Belgium: Centre for Research on the Economics of Development (CRED), Faculty of Economics, University of Namur. Retrieved from https://crespienrico.files.wordpress.com/2008/11/elite-capture.pdf

Plous, S. (1993). *The psychology of judgment and decision making.* New York: McGraw-Hill.

Plümper, T., Troeger, V. E., & Neumayer, E. (2010). Case selection and causal inference in qualitative research. *SSRN Electronic Journal.* Retrieved from http://dx.doi.org/10.2139/ssrn.1439868

Ravindra, A. (2004). *An assessment of the impact of Bangalore citizen report cards on the performance of public agencies.* Evaluation Capacity Development working paper series, No. ECD 12. Washington, DC: World Bank. Retrieved from http://siteresources.worldbank.org/INTEVACAPDEV/Resources/4585664-1253899870336/ecd_wp_12.pdf

Reinikka, R., & Svensson, J. (2004). *The power of information: Evidence from a newspaper campaign to reduce capture of public funds.* Washington, DC: World Bank. Retrieved from http://www1.worldbank.org/publicsector/decentralization/Feb2004Course/Background%20materials/Reinikka.pdf

Royal Netherlands Embassy. (2003). *Uganda: The fight against corruption. A case study on the prevalence of corruption in Uganda, specifically in local government, education, justice, law & order, and procurement.* Kampala, Uganda: Royal Netherlands Embassy.

Shankar, S. (2010). *Can social audits count?* ASARC Working Paper 2010/09. Canberra, Australia: The Australian National University, Australia South Asia Research Centre. Retrieved from https://crawford.anu.edu.au/acde/asarc/pdf/papers/2010/WP2010_09.pdf

Sheely, R. (2013). *Maintaining local public goods: Evidence from rural Kenya.* Cambridge, MA: Kennedy School of Government, Harvard University. Retrieved from http://scholar.harvard.edu/files/sheely/files/sheely-maintaining_local_public_goods-nov_2013.pdf

Social Impact, Solidaritas, and RTI International. (2013). *Kinerja impact assessment*. Prepared for USAID under the Kinerja program, Cooperative Agreement No. AID-497-A-10-00003. Arlington, VA: Social Impact.

Stein, D., & Valters, C. (2012). *Understanding theory of change in international development.* JSRP Paper 1. London, UK: London School of Economics. Retrieved from http://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/UNDERSTANDINGTHEORYOFChangeSteinValtersPN.pdf

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). *Broadening the range of designs and options for impact evaluations.* Working Paper 38. London, UK: Department for International Development. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf

Touchton, M., & Wampler, B. (2013). Improving social well-being through new democratic institutions. *Comparative Political Studies*, *47*(10), 1442–1469.

United States Agency for International Development (USAID). (2012). *USAID's global health strategic framework: Better health for development.* Washington, DC: USAID. Retrieved from https://www.usaid.gov/sites/default/files/documents/1864/gh_framework2012.pdf

USAID. (2013a). *USAID strategy on democracy, human rights and governance.* Washington, DC: Center of Excellence on Democracy, Human Rights and Governance, USAID. Retrieved from https://www.usaid.gov/sites/default/files/documents/1866/USAID%20DRG_%20final%20final%206-24%203%20(1).pdf

USAID. (2014a). *Evaluating the effectiveness of non-financial incentives to improve the delivery of health services in Sierra Leone* [Web page, Development Innovations Ventures]. Washington, DC: USAID. Retrieved from https://www.usaid.gov/div/portfolio/Health-Services-Sierra-Leone-IPA

USAID. (2014b). *User's guide to democracy, human rights, and governance programming.* Washington, DC: Center of Excellence on Democracy, Human Rights and Governance, USAID. Retrieved from https://www.usaid.gov/sites/default/files/documents/1866/DRGUsersGuide_Aug2014.pdf

Uphoff, N., & Wijayaratna, C. M. (2000). Demonstrated benefits from social capital: The productivity of farmer organizations in Gal Oya, Sri Lanka. *World Development*, *28*(11), 1875–1890.

Weller, N., & Barnes, J. (2014). Pathway analysis and the search for causal mechanisms. *Sociological Methods & Research.* http://dx.doi.org/10.1177/0049124114544420

Woolcock, M. (2009). Toward a plurality of methods in project evaluation: A contextualized approach to understanding impact trajectories and efficacy. *Journal of Development Effectiveness, 1*(1), 1–14.

World Bank. (2003). *World development report 2004: Making services work for poor people.* Washington, DC: International Bank for Reconstruction and Development / World Bank, and Oxford University Press. Retrieved from https://openknowledge.worldbank.org/handle/10986/5986

World Bank. (2007). *Social accountability sourcebook: Strengthening the demand side of governance and service delivery.* Washington, DC: World Bank. Retrieved from http://www.worldbank.org/socialaccountability_sourcebook/

ANNEX A. ILLUSTRATIVE INDICATORS FOR THE DESIGN STAGE

Type of indicator	Social accoun-tability aim	Illustrative input indicators	Illustrative output indicators
Illustrative Quality Indicators	Empowerment	social cohesion; awareness of entitlements; understanding of decision-making mechanisms (direct voting vs. representation); knowledge of the actions of Members of Parliament (MPs); knowledge of participatory mechanisms; knowledge of educational rights	social norms challenged (e.g., property rights restrictions and wage levels)
	Governance	use of citizen report cards; use of financial audits; the extent and enforcement of sanctions; decision-making mechanisms (direct voting vs. representation); knowledge of alternative dispute resolution mechanisms; political connectedness of health workers; leakage of funds mapped; audit irregularities found (corruption); manual of operations and school report cards; structured school meetings	democratic rules for selecting leaders; investment in monitoring institutions; the extent and enforcement of sanctions; reported transparency of leaders; likelihood of reelection; MP participation in Parliament
	Service Delivery	complexity of infrastructure improvements; the extent and enforcement of provider incentives and sanctions; reported transparency of service information; civil servant absence; service monitoring process	public perceptions of government services; maintenance quality; physical condition of infrastructure; citizen knowledge of/satisfaction of government actions; self-reported health status; reading skills
Illustrative Quantity Indicators	Empowerment	# of complaints; # of women in decision making positions; % of budget determined through participatory mechanisms; % of people who read newspapers; % of people who listen to radio stations	% of poor who access entitlements; % of poor children in school; % of budget spent on health and education; # of civil society organizations operating
	Governance	# of women in decision making positions; # of monitoring visits; wealth quintile of beneficiaries; # of complaints; % of leaders elected vs appointed; % of budget determined through participatory mechanisms; % of bureaucracies monitoring staff attendance; amount of municipal spending by sector and wealth quintile; estimated leakage of funds	% of incumbents reelected; # of monitoring reports received; # of audits published; % of complaints satisfactorily addressed; amount of public infrastructure investments; estimated leakage of funds; % of expenditures missing (wages and materials); % of budget spent on health and education; # of community disputes (e.g., land, water); # of community disputes resolved; voter turnout; % leakage of community grants; % of people observed not complying with government regulations

Type of indicator	Social accoun-tability aim	Illustrative input indicators	Illustrative output indicators
	Service Delivery	# of monitoring reports received; % of audit findings addressed and published; % of complaints satisfactorily addressed; amount of public infrastructure investment; % change in expenditures missing (wages and materials); % of budget spent on health and/or education; % change in leakage of community grants; civil servant absenteeism	# of hours of water/electricity per day; # of health clinic visits per person; # of years of education; % of poor who access entitlements; % of poor children in school; infant mortality rate; contraceptive prevalence rate; community service satisfaction; student attendance; grade repetition; dropout rates; exam pass rates

ANNEX B. ILLUSTRATIVE INDICATORS FOR THE IMPLEMENTATION STAGE

Type of indicator	Social accounta-bility aim	Illustrative input indicators	Illustrative output indicators
Illustrative Quality Indicators	Empowerment	social cohesion; awareness of entitlements; understanding of decision-making mechanisms (direct voting vs. representation); knowledge of the actions of Members of Parliament (MPs); know-ledge of participatory mechanisms; knowledge of educational rights	social norms challenged (e.g., property rights restrictions and wage levels)
	Governance	use of citizen report cards; use of financial audits; the extent and enforcement of sanctions; decision-making mechanisms (direct voting vs. representation); knowledge of alternative dispute-resolution mechanisms; political connectedness of health workers; leakage of funds mapped; audit irregularities found (corruption); manual of operations and school report cards; structured school meetings	democratic rules for selecting leaders; investment in monitoring institutions; the extent and enforcement of sanctions; reported transparency of leaders; likelihood of reelection; MP participation in Parliament
	Service Delivery	complexity of infrastructure improvements; the extent and enforcement of provider incentives and sanctions; reported transparency of service information; civil servant absence; service monitoring process	public perceptions of government services; maintenance quality; physical condition of infrastructure; citizen knowledge of/satisfaction of government actions; self-reported health status; reading skills
Illustrative Quantity Indicators	Empowerment	# of complaints; # of women in decision-making positions; % of budget determined through participatory mechanisms; % of people who read newspapers; % of people who listen to radio stations	% of poor who access entitlements; % of poor children in school; % of budget spent on health and education; # of civil society organizations operating
	Governance	# of women in decision-making positions; # of monitoring visits; wealth quintile of beneficiaries; # of complaints; % of leaders elected vs appointed; % of budget determined through participatory mechanisms; % of bureaucracies monitoring staff attendance; amount of municipal spending by sector and wealth quintile	% of incumbents reelected; % of complaints satisfactorily addressed; amount of public infrastructure investment; estimated leakage of funds; % of expenditures missing (wages and materials); % of budget spent on health and education; # of community disputes (e.g., land, water); # of community disputes resolved; voter turnout; % leakage of community grants; % of people observed not complying with government regulations
	Service Delivery	# of monitoring reports received; % of audit findings addressed; % of complaints satisfactorily addressed; amount of public infrastructure investment; % of expenditures missing (wages and materials); % of budget spent on health and/or education; % leakage of community grants; civil servant absenteeism	# of hours of water/electricity per day; # of health clinic visits per person; # of years of education; % of poor who access entitlements; % of poor children in school; infant mortality rate; contraceptive prevalence rate; student attendance; dropout rates; exam pass rates

ANNEX C. ILLUSTRATIVE INDICATORS FOR THE CLOSEOUT STAGE

Type of indicator	Social accountability aim	Illustrative input indicators	Illustrative output indicators
Illustrative Quality Indicators	Empowerment	social cohesion; awareness of entitlements; understanding of decision-making mechanisms (direct voting vs. representation); knowledge of the actions of Members of Parliament (MPs); knowledge of participatory mechanisms; knowledge of educational rights	social norms challenged (e.g., property rights restrictions and wage levels)
	Governance	use of citizen report cards; use of financial audits; the extent and enforcement of sanctions; decision-making mechanisms (direct voting vs. representation); knowledge of alternative dispute-resolution mechanisms; political connectedness of health workers; existence of manual of operations; existence of school report cards	democratic rules for selecting leaders; investment in monitoring institutions; the extent and enforcement of sanctions; MP participation in Parliament
	Service Delivery	complexity of infrastructure improvements; the extent and enforcement of provider incentives and sanctions; reported transparency of service information	public perceptions of government services; maintenance quality; physical condition of infrastructure; citizen knowledge of/satisfaction of government actions; self-reported health status; reading skills
Illustrative Quantity Indicators	Empowerment	# of women in decision-making positions; % of budget determined through participatory mechanisms; % of people who read newspapers; % of people who listen to radio stations	% of poor who access entitlements; % of poor children in school; % of budget spent on health and education; # of civil society organizations operating
	Governance	# of women in decision-making positions; wealth quintile of beneficiaries; % of budget determined through participatory mechanisms; % of bureaucracies monitoring staff attendance; amount of municipal spending by sector and wealth quintile	% of incumbents reelected; amount of public infrastructure investment; estimated leakage of funds; % of expenditures missing (wages and materials); % of budget spent on health and education; voter turnout
	Service Delivery	amount of public infrastructure investment; % of budget spent on health and/or education	# of hours of water/electricity per day; % of poor who access entitlements; % of poor children in school; infant mortality rate; contraceptive prevalence rate