Title

Pitfalls of “slippery indicators”: the importance of reading between the lines

Journal

Development in Practice, 33(6)

ISSN

0961-4524

Authors

Fischer-Mackey, Julia

Fox, Jonathan

Publication Date

2023-08-18

DOI

10.1080/09614524.2022.2104220

Development in Practice

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cdip20

Pitfalls of "slippery indicators": the importance of reading between the lines

Julia Fischer-Mackey & Jonathan Fox

To cite this article: Julia Fischer-Mackey & Jonathan Fox (2022): Pitfalls of "slippery indicators": the importance of reading between the lines, Development in Practice, DOI: 10.1080/09614524.2022.2104220

To link to this article: https://doi.org/10.1080/09614524.2022.2104220

Published online: 02 Aug 2022.

DEVELOPMENT IN PRACTICE
https://doi.org/10.1080/09614524.2022.2104220

Routledge
Taylor & Francis Group

Pitfalls of "slippery indicators": the importance of reading between the lines

Julia Fischer-Mackey and Jonathan Fox

ABSTRACT

Within the field of social accountability, studies about "community monitoring" have made broad claims about "what works" – or not – in practice, with significant implications for practitioners and policymakers. Interpretation of these findings is complicated when studies rely on "slippery indicators" that do not measure the real-world processes they claim to address. This article illustrates the problem of slippery indicators, which has two main elements. First, some studies rely on indicators that do not actually measure community monitoring. Second, studies that claim to show a failure of community monitoring to deliver improvements may actually show a failure to deliver community monitoring in the first place. While complex research methods may obscure these two related problems, readers can still assess whether studies' claims are supported by their empirical data by checking whether the findings are grounded in indicators that actually measure what they claim to study.

ARTICLE HISTORY
Received 28 March 2022
Accepted 17 July 2022

KEYWORDS
Accountability; participation; civil society; governance and public policy

Introduction

In recent years there has been a steady stream of large-scale studies of community monitoring and other social accountability interventions that result in null findings – with headlines suggesting these governance reform approaches "do not work".[1] For practitioners and policymakers, those findings can be puzzling – and indeed demoralising. It is true that many such efforts fail, but the problem with some of these studies is that they do not actually measure the real-world processes they intend to address. When researchers fail to define key concepts and choose indicators that are aligned with them, they may end up with "slippery indicators" that do not measure what they claim to measure. The problem of "slippery indicators", which this article discusses in the context of community monitoring studies, has two main elements: First, some studies draw conclusions about community monitoring, but they rely on indicators that do not measure community monitoring. Second, some studies that claim to observe the failure of community monitoring to deliver improvements in public services may actually show a failure to deliver community monitoring in the first place.

Although the question of valid measurement may sound like technical matter for methodologists, it actually matters a great deal to practitioners and policymakers seeking to learn what can be done to advance accountability. This article shows how non-specialist readers can assess the fine print of technical studies to assess whether they are grounded in convincing empirical indicators. This approach enables readers to assess the credibility of headline findings for themselves.

CONTACT Julia Fischer-Mackey juliafm@american.edu

The problem of slippery indicators

What do indicators really indicate? In social research and programme evaluation, the term "indicator" is used to mean a variable that is used to measure change related to a programme or policy objective. Indicators can be observed and measured, and they provide information about larger concepts that often cannot be directly observed or measured. A concept may refer to a process (such as community monitoring) or a phenomenon (such as budget transparency). Because measuring processes of change in relationships between citizens and the state is difficult, multiple empirical indicators are often needed to provide meaningful insights into governance issues.

A technical term for assessing "whether a variable measures what it is supposed to measure" is *measurement validity*.[2] A *valid measure* depends on two things. First, it requires an explicit definition of the concept of interest (Adcock and Collier 2001; Sartori 1984).[3] This is crucial because key concepts of interest to social scientists can be defined in many different ways. Many concepts in the governance field are ambiguous and contested – including participation, empowerment, community monitoring and accountability (e.g. Alsop, Bertelson, and Holland 2006; Fox 2022; Narayan 2005; Shutt and McGee 2013; VeneKlasen and Miller 2007). In the context of multiple possible definitions and approaches to translating these concepts into measurable empirical indicators, researchers need to spell out the rationale for their choices of empirical indicators.

The second requirement for a *valid measure* is that the empirical indicators both capture core elements of the concept and do not include a lot of unrelated information. Simply put, the indicators must tell us something meaningful about the existence of a key concept or a change in it. To explain the rationale for their choices of indicators, researchers need to address at least two related issues. First, how does an indicator capture a concept? Second, what are the assumptions, strengths, and limitations of that indicator? If the indicator's relationship to the concept is clear, logical, or justified by the researchers, the reader can feel confident in it. While some researchers may be comfortable measuring what they assume to be an indirect indicator of the process of interest, others may not be convinced by those assumptions. If the indicator's relationship to the concept is unclear or very indirect, the researchers owe an explanation to readers about why the indicator is appropriate to use.

If researchers fail to define their key concepts or to explain their rationale for selecting specific indicators, they may end up with what we call here "slippery indicators". Like a slippery substance that is hard to get a handle on, slippery indicators make it difficult for readers to know what they really say about the key concept. The problem of slippery indicators can be observed in many types of research. To illustrate the problem, we discuss examples primarily from the experimental literature. Even though field experiments may use rigorous sampling and analytical techniques, the validity of their findings depends on whether their underlying empirical evidence is grounded in convincing measures of clearly defined concepts. In other words, technically rigorous data analysis does not address whether the process of interest was measured convincingly in the first place.[4] This is where non-technical readers can draw their own conclusions about the soundness of the study, even if they not have the technical training to assess the randomisation, sampling techniques, or quantitative analysis.

Slippery indicators and unspecified concepts also contribute to confusion about how to interpret empirical research findings. Practitioners want to know: "what's the takeaway?" Yet making sense of complex empirical findings is often more a matter of interpretation and spelling out the logic used than a technical methodological issue. This is a major issue with research in the broad field of governance reform because some studies claim that interventions fail to deliver the expected improvements in service delivery, when the evidence actually presented shows that the interventions were not delivered in the first place.[5] This issue of interpretation comes up in studies of community monitoring, when attempts to induce community monitoring are treated as evidence that community monitoring actually happened. Since attempts to encourage community monitoring may or not

work, researchers need to provide evidence that community monitoring *actually occurred* if they are going to claim that community monitoring did not "work" to improve service provision.[6]

The concept of community monitoring within the "social accountability" field

Social accountability is an umbrella category that includes numerous "formal or informal mechanisms through which citizens and/or civil society organisations engage to bring state officials or service providers to account" (Camargo and Jacobs 2013, 7). Both practitioner and scholarly interest in social accountability grew quickly after the 2004 *World Development Report* recognised that citizen voice could directly contribute to improved service provision (World Bank 2004). This "short route" to accountability would be, in theory, more efficient at improving services than the lengthy process of electoral accountability and subsequent managerial efforts.[7]

Social accountability efforts often involve some type of information provision or access and the monitoring of public services or actions, to inform citizen action. Some approaches to monitoring rely on increasing the transparency or dissemination of official data, and some involve data collection and analysis by technically sophisticated NGOs. Another approach to information production and dissemination involves "community monitoring", which is the collection of information about public services *by members of the community* those services are meant to reach, in order to document problems and possibly to track improvements. Community monitoring activities may take many forms, but they always involve inclusive processes of generating and analysing data. As Joshi explains: "Often, community monitoring is used as a way of ensuring that ongoing performance maintains normal standards and is focused on observable features, for example, teacher or doctor attendance, quality of construction in facilities or appropriate procedures being followed" (2013, S37).

This definition distinguishes between collecting data and making sense of it (the "community monitoring"), and possible follow-on "community actions" that are informed by the data. Such actions may include drawing public or media attention to an issue, requesting information from authorities, collaborative problem-solving with public officials, holding authorities accountable through administrative recourse, advocating to elected officials, or by using the legal system or social sanctioning (e.g. Hernández et al. 2019; Fischer-Mackey et al. 2020; Guerzovich and Poli 2020). It is important to note that community action to improve service provision may also occur *without* community monitoring; likewise, data collection and analysis may occur *without* subsequent collective action. Therefore, to conflate the concept of community monitoring with subsequent voice and action would complicate efforts to identify its potential contribution.

Information that is generated by community members themselves may be more credible to them and more relevant to their needs (Flores 2018). Community monitoring data can also legitimate community members' concerns in the eyes of service providers and public officials (even if they document problems that were already locally known, such as medicine stockouts or patterns of disrespect and abuse).[8]

Other authors' conceptualisations of community monitoring are much broader. For example, a systematic review includes external information campaigns that NGOs or others use to *encourage* community members to take action as a *type* of community monitoring (Molina et al. 2017, 5–6). However, efforts to provide information or encouragement to communities are possible precursors that may or may not lead to *actual* monitoring by communities. The same evidence review also includes grievance redress mechanisms (GRMs) as a *type* of community monitoring. GRMs are institutionalised processes that allow citizens to report complaints to governments, to seek redress for services they rightfully should have received (Pande and Hossain 2022). GRMs do not involve community data collection and their findings are rarely made public. They provide information – but to the government, not to communities. The point here is that community monitoring refers to one distinct approach to the production of evidence, within a broad constellation of social accountability strategies and tactics. To define all tools for local information production or provision as community

monitoring, or to treat the concept as interchangeable with the broad umbrella category of social accountability, complicates efforts to identify the strengths and limitations that are *specific* to community monitoring.

To summarise, community monitoring may take many forms, but at its core, the concept is about the collection of information about public services *by members of an affected community* for the purposes of documenting problems and supporting actions to address them. Here follow brief examples of studies that make claims about community monitoring without actually measuring such actions.

Slippery community monitoring indicators

"Power to the People" (Björkman and Svensson 2009) was a very influential field experiment that provided communities in Uganda with report cards containing data on health centre utilisation and perceptions of service quality. The study intended to assess whether information about health services would spur people to act in ways that would lead to improved health services and health outcomes. The study did find improved health outcomes and attributed them to "community monitoring". Although the study did not define what it meant by "community monitoring", it operationalised the concept with an index including six indicators (see Box 1). Of the six indicators, only one was related to the concept of community monitoring as discussed above.

Box 1: Slippery "community monitoring" indicators in Björkman and Svensson (2009, 749, Table 2**)**

(1) Suggestion boxes
(2) Numbered waiting cards
(3) Poster informing free services
(4) Poster on patients' rights
(5) **Discuss facility in [Local Council] meetings**
(6) Received information about [Health Unit Management Committee]

This index does not reflect any coherent definition of community monitoring, and it conflates loosely related and unrelated processes with community monitoring. The authors do not explain why these indicators should be taken as evidence of community monitoring processes or outcomes. Three of the six indicators (#3, #4, #6) represent one-way information dissemination *to* communities, but not about clinic performance. Indicators #1 and #2 represent tools that provide information *to clinic managers*. The study does not specify why or how they may be related to community data collection or use. Only Indicator #5 (whether community members discussed the health facility in community meetings) could possibly indicate community monitoring – but even then, the justification for this assumption is not explained. In spite of this dilution of actual community monitoring in the index that is the basis for the quantitative analysis, the study also includes additional evidence – not included in the index – that is a much more plausible explanation of improved service provision. Before the intervention, the official health facility management committees were widely viewed as ineffective. In the treatment areas, one third of those committees were reconstituted or elected new leadership, in contrast to no such changes in the control group (Björkman and Svensson 2009, 747). These leadership changes in official bodies charged to do oversight are convincing indicators that community monitoring was invigorated. In contrast, reports of posters are indicators of informational campaigns (as noted above) rather than community monitoring.

A related follow-up study to "Power to the People" by Raffler and colleagues (2019a, 2019b, 2020), attempted to replicate the original study's general findings on a larger scale. However, they did not find positive impacts on health outcomes.[9] Raffler and colleagues stated that they documented what they called community monitoring, but they included few indicators of community monitoring. Only one in four components of the Raffler, Posner, and Parkerson (2019a) index may have been a reasonable indicator of community monitoring – a question about whether the clinic was discussed in community meetings (see Box 2 below).

Box 2: Slippery "community monitoring" indicators in Raffler, Posner, and Parkerson (2020, A3)

(7) "Whether household members report having attended [Local Council] meetings in the last year"
(8) **"Whether household members who attended [Local Council] meeting report that local health centre was discussed"**
(9) "Whether household members *think* engaged community members would find out if a health worker did not provide the effort that he/she should in caring for his/her patients" (emphasis added)
(10) "Whether household members *think* engaged community members would find out if a health worker did not report for work" (emphasis added)

As shown in Box 2, one of the four questions in the study's community monitoring index (#1) simply asked about attendance at community meetings (a weak indicator of participation, but not of monitoring). Two others (#3 & #4) involved questions that ask only about hypothetical scenarios rather than about whether community monitoring actually happened and, if so, whether and how service providers responded (2019a, Annex A, 40). A much more direct measure of community monitoring would have addressed whether they actually documented absenteeism. The study does not explain the rationale for its reliance on very indirect indicators that conflate hypothetical questions about perceived agency with evidence of actual community monitoring.

The authors state that "We also find no evidence that citizens increased their monitoring or sanctioning of health workers" (Raffler, Posner, and Parkerson 2020, title page). Yet the data presented does not include a direct measure of community monitoring, so it is unclear how the authors would know if it occurred. Moreover, in spite of their interpretation that their intervention did not produce citizen pressure, they go on to claim that "Contra the literature, the link between information provision and [health] provider behaviour did not run through citizen pressure" (Raffler, Posner, and Parkerson 2020, 25); in other words, citizen pressure does not lead to improved services. They do not explain the rationale for their conclusion that citizen pressure fails to influence service provision, while at the same saying they have no evidence that community monitoring even occurred. Notably, after the RCT concluded and the constraints of the study protocols were lifted, the same donor project supported a health rights campaign that used community monitoring of health services to inform multi-level advocacy, while also monitoring government responsiveness (Bailey and Mujune 2021). Both the original Björkman and Svensson (2009) and follow-up Raffler, Posner, and Parkerson (2019a; 2020) studies fail to define community monitoring, include weak indicators of it, and complicate the interpretation of their data by including measures of unrelated phenomena in their indices.

Another notable example of slippery indicators is in the education sector: "Pitfalls of participatory programs: Evidence from a randomised evaluation in education in India" (Banerjee et al. 2010). The study set out to test whether "participation of beneficiaries in the monitoring of public services" (Banerjee et al. 2010, 1) could inform action to improve educational outcomes. Yet the study found no evidence that monitoring of schools occurred. Instead, the intervention included monitoring of student-level learning outcomes and an unsuccessful attempt to activate official school oversight committees.

First, community volunteers collected data on student-level learning by giving assessments to individual children, finding low levels of learning. However, poor learning outcomes may be attributed to a variety of social, economic, personal, and institutional problems (e.g. social and emotional problems, learning disabilities, low school attendance, hunger and stress, mismatch between language of instruction and language spoken at home, teacher absenteeism, low quality instruction, etc.) Therefore, student-level learning data is not a substitute for data about public school services such as the school facilities and supplies, textbooks, teacher quality, teacher attendance, curriculum, or other indicators that directly relate to educational services provided. In addition, the process of community members collecting and analysing data about the public services can be an important step in developing a plan to address problems. Therefore, the individual nature of volunteers

testing children is fundamentally different from the core concept of community monitoring of service provision.

Second, the study was premised on the idea that the information on poor student learning outcomes would lead community members and official educational oversight bodies (Village Education Committees, or VECs) to improve the quality of the education provided in village schools. VECs are local appointed bodies composed of the head teacher, head of village government, and three parents chosen by government officials. In principle, the VECs could monitor teachers, complain about specific teachers to education authorities, hire additional assistant teachers, allocate minimal resources, and petition the government for more resources. The only monitoring of public services that might have taken place under the design of the program would have been by the VECs. Yet the study did not indicate whether relevant school-level performance data (such as teacher absenteeism or textbook availability) was available to the VECs or the community.

Before the intervention, a vast baseline study found that the VECs existed only on paper (Banerjee et al. 2010, 13–14). However, even after the interventions intended to activate the oversight committees, only seven percent of parents reported that they knew the VECs existed (Banerjee et al. 2010, 22). This, along with the absence of any related indicators, suggests that the VECs were not actively involved in monitoring the schools.

The study's point of departure was that "beneficiaries" must "have the necessary information to monitor the providers" (Banerjee et al. 2010, 2). Yet its "community monitoring" focused instead on individual student learning. Therefore, it is unclear whether community members in fact blamed poor student learning on the school system or on the students themselves, poverty, or on some other factor. The article's title – "Pitfalls of participatory programs" – suggests that the study tested a program that was participatory but failed to deliver the expected improvements in service delivery. Instead, the study documented something else: an intervention that failed to induce community monitoring in the first place, and thus did not test the impact of participatory monitoring on service delivery.

Better community monitoring indicators

An example of strong community monitoring measures can be found in a study of citizen-led initiatives for the right to health in Guatemala (Hernández et al. 2019). The study reported on two types of community monitoring. The first was that data was collected by citizens about health facilities' services. For the purposes of a Qualitative Comparative Analysis, the monitoring of health facilities was scored according to the: "[l]evel of participation in and continuity of collection of evidence through health facility visits, user interviews, community assemblies, [and] SMS complaint platform", with the highest score being assigned to examples where "Regular monitoring activities [occurred] for more than 1 year, generating large amount of evidence" (Hernández et al. 2019, 395).

The second type of monitoring was about the engagement with authorities by volunteer defenders of the right to health and NGO partners, which included the progress being made towards citizens' desired outcomes. The engagement with authorities was assessed based on

> Frequency and focus of meetings with municipal, district health, and higher level authorities, including provincial and national level authorities, and human rights institutions; signs of follow up with and active support (e.g. provision of meeting space) from municipal, district health, and higher level authorities; frequent interaction with municipal authorities with focus on action and follow up on problems presented; and frequent interaction with district health authorities with focus on action and follow up on problems presented; and audience with 2 or more higher level authorities or multiple interactions to follow up on problem presented. (Hernández et al. 2019, 395)

These empirical indicators of community monitoring document the community monitoring that *actually* occurred, not simply that efforts were undertaken to promote the monitoring. In addition, it documents monitoring of follow-on citizen action that is informed by the monitoring data – which can then be analysed to inform refined social accountability strategies. This example demonstrates

how to avoid the problem of slippery indicators by providing readers with a precise understanding of what the study actually observed and of the evidence for its conclusions.

Discussion and conclusions

This paper illustrates the problem of "slippery indicators" with analysis of studies that analyse community monitoring to improve public service provision. This new concept of slippery indicators names a disconnect between what researchers' headline findings claim and what their measures actually measure.

This discussion indicates that even notable field experiments that are rigorous in their randomisation, sample selection, and analysis of data may not be grounded in convincing empirical measures of the key concepts they study. While practitioners and policymakers may not be positioned to assess the technical side of data analysis, they are well-positioned to assess whether researchers make convincing assumptions about their indicators. For readers to make informed decisions about how to assess research findings, researchers need to be more consistent and explicit about how they define key concepts – and the rationale for how they are measured empirically.

Busy readers may assume that academic journals' external review processes assure that the researchers' interpretations are aligned with their actual empirical findings. The problem of slippery indicators – at least in the fields of governance and sectoral reform – suggests that readers cannot take such consistency for granted. The patterns of disconnects illustrated here, where headlines and summaries turn out not to be grounded in the evidence, suggest that it is worth a reader's time to review the authors' definitions of concepts and empirical indicators and ask – how were those research choices justified? Do the indicators actually indicate what they claim? This article has identified several cases where headline findings made unexamined assumptions about how to interpret the underlying data.

The question of whether indicators really indicate what researchers claim comes front and centre when it comes to claims about "what works". In the studies of governance and sectoral reform, authors' conclusions, abstracts and article titles often deploy authoritative messages framed in terms of whether specific kinds of transparency, participation and accountability interventions were found to "work" – or not. Yet this discussion shows that researcher assessments of "what works" can be a matter of interpretation – especially if the interventions encountered pitfalls in their design or implementation.

For practitioners, *how* to deliver governance reform is a central concern – yet the "does it work" framing draws researchers away from the challenge of *how* to pull off governance reform in the first place. To be convincing, studies that claim to assess whether social accountability "works" need to show that those processes actually happened in practice. To sum up this issue of how to interpret empirical findings, for practitioners as well as analysts, there is a big difference between *governance reforms that fail to deliver* – and *the failure to deliver governance reforms*. The first place for informed readers to look out for this kind of slippage between headlines and findings is in the empirical indicators.

In conclusion, the main contribution of this paper is to introduce the concept of slippery indicators – and to encourage non-specialist readers to look out for them in order to assess the credibility of research they encounter.

Notes

1. These studies focus on interventions that leverage the power of information to stimulate and inform civic action to improve public sector performance. See, among others, Banerjee et al. (2010); Björkman and Svensson (2009); Raffler, Posner, and Parkerson (2019a, 2019b, 2020); Kosec and Wantchekon (2020); Lieberman, Posner, and Tsai (2014).
2. See Bollen (1989, 184) and King, Keohane, and Verba (1994, 25), cited in Adcock and Collier (2001, 530).

3. It is also important that indicators do not capture processes that are unrelated to the concept of interest. According to Sartori, clear concepts allow analysts to "distinguish A from whatever is not-A" (1984, 74).
4. This issue of whether the indicators capture the phenomenon of interest is distinct from the question of whether an intervention produced null results because it was too weak to make a different – the "low dose" problem. See Fox (2015) and Wein (2020)
5. For one example of the distance between headline findings and underlying evidence, a recent major multi-country study claimed that community policing failed to deliver, when the evidence presented shows instead a failure to deliver community policing (Blair 2021). Authors of a field experiment involving citizen training to deepen local democracy declared failure when its ethnographic evidence showed that in practice the intervention was incompletely delivered (Rao, Ananthpur, and Malik 2017). A multi-country test of the power of information to encourage electoral action also declared null results, yet the interventions were weak and incomplete delivered (Dunning et al. 2019). For another example, a study that claimed to find that a transparency and accountability intervention failed to deliver, actually measured the effects of a participation intervention involving little transparency or accountability (Arkedis et al. 2021).
6. Authors of one field experiment that assesses the effect of participation on accountability explicitly claims that whether citizen actions for accountability actually take place is not the point, suggesting that the potential threat of such action can be equally effective. In this view, to provide relevant information is assumed to be as powerful as if citizens actually took action: " … since both an actual protest and the threat of voice may have discouraged the local political elite from diverting resources intended for the schools, in equilibrium, there is no reason to believe that the incidence of voice and local diversion of funds should be correlated" (Reinikka and Svensson 2011, 959). The proposition is treated as self-evident rather than as a hypothesis that needed an explicit rationale and testing.
7. Subsequent analysis distinguished between responses and responsiveness to citizen voice. In practice, official responses to citizen voice may fall short of responsiveness – as when improvements in service provision turn out to be one-off, partial or selective (Fox 2022).
8. The kind of monitoring results that generate credibility with local authorities may be different from the kind of evidence intended to meet social science standards. Flores (2018) describes how grassroots community monitors collected user testimony instead of conventional quantitative data on health system problems because it was more influential with local authorities.
9. The public draft of the study does not report the specific components of the report card, beyond general references to utilisation and perceptions of quality and satisfaction. The perceived relevance of these indicators to citizens is not discussed. The report cards apparently did not include more specific, tangible indicators of clinic performance, such as medicine stockout rates, staff absenteeism, or unfilled positions.

Acknowledgements

Thanks very much to Angela Bailey for the extensive discussions that informed this article. Thanks to the anonymous reviewers, as well as to Adam Auerbach, Robert Adcock, Susanna Campbell, and Rachel Sullivan Robinson for their comments on earlier drafts.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Julia Fischer-Mackey, Ph.D., is a lecturer at American University's School of International Service.

Jonathan Fox is a professor of development studies at the School of International Service at American University, where he directs the Accountability Research Center (www.accountabilityresearch.org).

ORCID

Julia Fischer-Mackey http://orcid.org/0000-0002-0336-3283
Jonathan Fox http://orcid.org/0000-0002-2889-0397

References

Adcock, R., and D. Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95 (3): 529–546.

Alsop, R., M. Bertelson, and J. Holland. 2006. *Empowerment in Practice: From Analysis to Implementation*. Washington, DC: World Bank.

Arkedis, J., J. Creighton, A. Dixit, A. Fung, S. Kosack, D. Levy, and C. Tolmie. 2021. "Can Transparency and Accountability Programs Improve Health? Experimental Evidence from Indonesia and Tanzania." *World Development* 142: 105369.

Bailey, A., and V. Mujune. 2021. *Bottom-Up Accountability in Uganda: Learning from People-Centered, Multi-Level Health Advocacy Campaigns*. Working Paper No. 8. Washington, DC: Accountability Research Center. https://accountabilityresearch.org/publication/bottom-up-accountability-in-uganda-learning-from-people-centered-multi-level-health-advocacy-campaigns/.

Banerjee, A. V., R. Banerji, E. Duflo, R. Glennerster, and S. Khemani. 2010. "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India." *American Economic Journal: Economic Policy* 2 (1): 1–30.

Björkman, M., and J. Svensson. 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.

Blair, G. 2021. "Community Policing Does Not Build Citizen Trust in Police or Reduce Crime in the Global South." *Science* 374 (6571): 1–14.

Bollen, K. A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Camargo, C. B., and E. Jacobs. 2013. *Social Accountability and its Conceptual Challenges: An Analytical Framework*. Working Paper 16. Basel: Basel Institute on Governance. https://edoc.unibas.ch/66317/.

Dunning, T., G. Grossman, M. Humphreys, S. D. Hyde, C. McIntosh, and G. Nellis. 2019. *Information, Accountability and Cumulative Learning: Lessons from Metaketa I*. Cambridge: Cambridge University Press.

Fischer-Mackey, J., B. Batzin, P. Culum, and J. Fox. 2020. "Rural Public Health Systems and Accountability Politics: Insights from Grassroots Health Rights Defenders in Guatemala." *Journal of Peasant Studies* 47 (5): 899–926.

Flores, W. 2018. *How Can Evidence Bolster Citizen Action? Learning and Adapting for Accountable Public Health in Guatemala*. Accountability Note 2. Washington, DC: Accountability Research Center. https://accountabilityresearch.org/publication/how-can-evidence-bolster-citizen-action-learning-and-adapting-for-accountable-public-health-in-guatemala/.

Fox, J. 2015. "Social Accountability: What Does the Evidence Really say?" *World Development* 72: 346–361.

Fox, J. 2022. *Accountability Keywords*. Working Paper 11. Washington, DC: Accountability Research Center. https://accountabilityresearch.org/publication/accountability-keywords/.

Guerzovich, M. F., and M. Poli. 2020. *How Social Accountability Strengthens Cross-Sector Initiatives to Deliver Quality Health Services? Global Partnership for Social Accountability Note 17*. Washington, DC: The World Bank. https://documents1.worldbank.org/curated/en/600891606911830725/pdf/How-Social-Accountability-Strengthens-Cross-Sector-Initiatives-to-Deliver-Quality-Health-Services.pdf.

Hernández, A., A. L. Ruano, A. Hurtig, I. Goicolea, M. San Sebastián, and W. Flores. 2019. "Pathways to Accountability in Rural Guatemala: A Qualitative Comparative Analysis of Citizen-Led Initiatives for the Right to Health of Indigenous Populations." *World Development* 113: 392–401.

Joshi, A. 2013. "Do They Work? Assessing the Impact of Transparency and Accountability Initiatives in Service Delivery." *Development Policy Review* 31: s29–s48.

King, G., R. O. Keohane, and S. Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

Kosec, K., and L. Wantchekon. 2020. "Can Information Improve Rural Governance and Service Delivery?" *World Development* 125: 1–13.

Lieberman, E., D. Posner, and L. Tsai. 2014. "Does Information Lead to More Active Citizenship? Evidence from an Education Intervention in Rural Kenya." *World Development* 60: 69–83.

Molina, E., L. Carella, A. Pacheco, G. Cruces, and L. Gasparini. 2017. "Community Monitoring Interventions to Curb Corruption and Increase Access and Quality in Service Delivery: A Systematic Review." *Journal of Development Effectiveness* 9 (4): 462–499.

Narayan, D. 2005. *Measuring Empowerment: Cross-Disciplinary Perspectives*. Washington, DC: The World Bank.

Pande, S., and N. Hossain. 2022. *Grievance Redress Mechanisms in the Public Sector A Literature Review*. Washington, DC: Open Governance Partnership and Accountability Research Center. https://www.opengovpartnership.org/documents/grievance-redress-mechanisms-in-the-public-sector-a-literature-review/.

Raffler, P., D. Posner, and D. Parkerson. 2019a. "The Weakness of Bottom-Up Accountability: Experimental Evidence from the Ugandan Health Sector." *Innovations for Poverty Action* (March 26 Draft). https://www.poverty-action.org/sites/default/files/publications/The%20Weakness%20of%20Bottom-Up%20Accountability.pdf.

Raffler, P., D. Posner, and D. Parkerson. 2019b. "Does Bottom-Up Accountability Work? Evidence from Uganda." *VoxDev*. July 22. https://voxdev.org/topic/health-education/does-bottom-accountability-work-evidence-uganda.

Raffler, P., D. Posner, and D. Parkerson. 2020. "Can Citizen Pressure be Induced to Improve Public Service Provision?" *Innovations for Poverty Action* (Oct. 30 draft). https://www.poverty-action.org/sites/default/files/publications/Citizen%20Pressure_Oct302020.pdf.

Rao, V., K. Ananthpur, and K. Malik. 2017. "The Anatomy of Failure: An Ethnography of a Randomized Trial to Deepen Democracy in Rural India." *World Development* 99: 481–497.

Reinikka, R., and J. Svensson. 2011. "The Power of Information in Public Services: Evidence from Education in Uganda." *Journal of Public Economics* 95: 956–966.

Sartori, G. 1984. *Social Science Concepts: A Systematic Analysis*. Thousand Oaks, CA: Sage.

Shutt, C., and R. McGee. 2013. *Improving the Evaluability of INGO Empowerment and Accountability Programmes*. Practice Paper. No. 1. Sussex: Centre for Development Impact. https://www.ids.ac.uk/publications/improving-the-evaluability-of-ingo-empowerment-and-accountability-programmes/.

VeneKlasen, L., and V. Miller, eds. 2007. *A New Weave of Power, People and Politics: The Action Guide for Advocacy and Citizen Participation*. Rugby: Practical Action Publishing.

Wein, T. 2020. "Intervening Intensely: How to Estimate Dosage in Accountability Interventions." April 23.. https://medium.com/@tom_wein/intervening-intensely-how-toestimate-dosage-in-accountability-interventions-1da1b51d051a.

World Bank. 2004. *World Development Report 2004: Making Services Work for Poor People*. Washington, DC: World Bank.