Encouraging Participation

Discussion draft; comments welcome

This draft: December 5, 2019

Stephen Kosack (University of Washington, Seattle)

Grant Bridgman (Uliza)

Jessica Creighton (Harvard University)

Courtney Tolmie (Results for Development Institute)

Preston Whitt (Results for Development Institute)

Archon Fung (Harvard University)

Abstract: Can civic engagement that is encouraged by a development program be empowering and helpful for improving public services? Transparency and accountability or social accountability programs are a popular approach to improving the responsiveness and effectiveness of health care, education, and other public services, but evidence of their effects is mixed. We ask whether participants in 200 randomly selected communities in Indonesia and Tanzania engaged with an experimental community-led scorecard program they were offered and whether they found the experience to be empowering and helpful for improving their maternal and newborn health care. Interviews, focus groups, and observations of program meetings before, during, immediately after, and two years after the program all indicate from complementary perspectives that in almost all communities, participants engaged in sustained and largely self-directed discussions about how to improve their care and tried the approaches they designed. Although their experiences were far from uniform and some grew skeptical of their efficacy, most who participated throughout eventually described their activities as having improved their care and as being as or more confident in their capacities to improve their communities than when they began. On average, their efforts were not sufficient to add significantly to measurable health outcomes in their broader communities two years after the program, relative to 200 other communities who were not offered the program. Thus participants' perceptions of the efficacy of their efforts might stem in part from attribution bias. Yet the evidence also suggests that in a substantial minority of communities, participants had been willing to continue their efforts long after the program ended, and that their efforts had led to changes in their care that, although they may have had limited effects on average community-level outcomes or substituted for others' efforts, were noticeable and memorable to them and others in their community. Altogether a wide range of observations and reflections all suggest that for most who participated throughout the program, the experience sustained or improved their perceptions of civic efficacy.

Acknowledgements: This paper stems from the Transparency for Development project, a multi-country mixed method research project that seeks generalizable knowledge about the impacts and implications of transparency and accountability programs. It builds from six years of work with colleagues at PATTIRO in Indonesia, CHAI in Tanzania, the Results for Development Institute, the Evans School of Public Policy and Governance at the University of Washington, and the Ash Center for Democratic Governance and Innovation at the Harvard Kennedy School. Over the years we have drawn insight, ideas, and inspiration from many scholars and practitioners in the fields of transparency and accountability and social accountability and related disciplines, among them Yamini Aiyer, Rikhil Bhavnani, Kevin Bohrer, Nick Burnett, Jessica Cohen, Jishnu Das, Jonathan Fox, Rachel Glennerster, Mary Graham, Florencia Guerzovich, Scott Guggenheim, Hahrie Han, Patrick Heller, Vanessa Herringshaw, Michael Jarvis, Anuradha Joshi, Asim Khwaja, Jenny Lah, Jane Mansbridge, Scott Mainwaring, Tom McEnroe, Jim McGuire, Nushina Mir, Ben Olken, Rohini Pande, Alfonsina Peñaloza, Enrique Peruzzotti, Dan Posner, Lant Pritchett, the late Helen Richards, Susan Rose-Ackerman, Bill Savedoff, Jakob Svensson, Lily Tsai, Michael Walton, David Weil, Tom Wingfield, and Michael Woolcock, in addition to current and former members of the Transparency for Development research team: Sarah Alphs, Iqra Anugrah, Novi Anggraeni, Jean Arkedis, Megan Cogburn, Sad Dien, Akshay Dixit, Riska Hasan, Hannah Hilligoss, Jenna Juwono, Dan Levy, Hilda Kigola, Esther Mtumbuka, Rohit Naimpally, Happy Ndomba, Didik Purwandanu, Mohamed Yunus Rafiq, James Rasaiah, Lindsey Roots, Rachmat Reksa Samudra, Manpreet Singh, Astri Arini Waluyo, Preston Whitt, and Xie Kankan. Our thanks as well to Wayan Suriastini and her team at SurveyMeter in Indonesia and Samson Kiware, Isaac Lyatuu, Diego Shirima and the rest of their team at Ideas in Action in Tanzania for superb data collection. The study was made possible by generous financial support from the Bill and Melinda Gates Foundation, the William and Flora Hewlett Foundation, and the U.K. Department for International Development, under the coordination of the Transparency and Accountability Initiative. Most importantly, we gratefully acknowledge the thousands of Indonesians and Tanzanians in the communities who participated in the program. More on the project, including background documents and reports and other working papers, is available at t4d.ash.harvard.edu.

1. Introduction

Following a decades-long worldwide commitment to supporting basic human capabilities, modern health care, among other public services, has expanded rapidly across the world (UNDP 1990; Sen 1999; UN General Assembly 2000; World Bank 2004). Yet despite substantial advances in access and quality, many communities still lack access to clean, well-equipped and -staffed clinics offering responsive and accessible care, and further improvements remain at the core of the next generation of international commitments to sustainable development (UN General Assembly 2015).[1] Alongside commitments of resources, equipment, financing, and other assistance, nongovernmental organizations and international donors have turned increasingly in recent decades to an approach to improving the responsiveness and effectiveness of public services like health care, known loosely as "transparency and accountability" or "social accountability."[2]

When the workings of public services are transparent to the communities they serve, those who use these services might be in an advantageous position to improve them.[3] They might work with staff and officials to improve facilities or supplies or how patients or students are treated, complain to providers when they are dissatisfied with services, organize themselves to improve schools or clinics, ask officials or political representatives for help, post hours and charges, help each other understand when and how to use the service, create transportation pools for accessing hard to reach services, and countless other possibilities. The typical transparency and accountability or social accountability program offers citizens or civic organizations information on the workings of public services—test scores or health metrics, evidence of provider attendance, supply stocks, budget data—and often discussion and decision-making spaces as well to encourage them to act or convince others to act in ways that improve the responsiveness and effectiveness of those services.

Can programs like these encourage participation that is empowering and helpful for improving public services? A number of studies find that people who participate in a transparency and accountability program can make measurable, and sometimes transformative, improvements to their public services (e.g. Björkman and Svensson 2009; World Bank 2018; Banerjee et al. 2018).[4] But other studies find the opposite

[1] See also Kruk et al. (2016), World Economic Forum (2015), Farmer et al. (2013), Hsia et al. (2012). The recent *Lancet* Global Health Commission on High Quality Health Systems in the SDG era (Kruk et al. 2018) defines quality health care in line with the SDGs as "informed by four values: they are for people, and they are equitable, resilient, and efficient" and conclude that high quality health systems in which patients have trust and confidence could save 8 million lives each year in low and middle income countries.

[2] World Development Report (2004), Fox (2007a; 2015), Joshi (2010), Gaventa and McGee (2011), J-PAL (2011), Kosack and Fung (2014). One example of the growth of such programs is the UK Department for International Development's "empowerment and accountability" portfolio, which at the time of a 2016 review had 2,379 projects, including 180 in its 28 priority countries that used a range of approaches to encouraging social accountability (Holland and Schatz 2016, 16).

[3] Much of the international effort toward improving public services is committed to seeking improvements in public services in participation and partnership with the communities they are designed to serve (OECD 2008; UN General Assembly 2015).

[4] There have been several dozen randomized controlled trials of transparency and accountability or social accountability programs; see reviews in: Gaventa and McGee (2011), J-PAL (2011), Kosack and Fung (2014), Fox (2015).

(e.g. Banerjee et al. 2010; Ananthpur, Malik, and Rao 2014; Raffler, Posner, and Parkerson 2019), and many question the sufficiency of programs that do not also include additional resources, authority, or some other source of empowering support (Baiocchi, Heller, and Silva 2011; Gaventa 2006; Fung 2006; Fung and Wright 2003) or when those services are not supplemented (Sachs 2005) or the state providing them is not also supported with additional resources or capacity with which to respond (Andrews, Pritchett, and Woolcock 2017; Fox 2015; Besley and Persson 2010; Grindle 1997; 2007).

Underlying the approach are several assumptions, two of which we concentrate on in this paper: that in the communities where transparency and accountability programs are offered there are some people who 1) are capable of acting or convincing others to act in ways that could lead to improvements in their public services, and who 2) would try if they were encouraged to as part of a transparency and accountability program. Even where these assumptions hold, the relationships between transparency and participation and more responsive and effective public services will be complex and contextually contingent.[5] Yet even these basic assumptions may not hold in many communities. Many may not include people who are willing and able to take time from family and work to try to improve a public service when they are encouraged to; many might instead ignore or resist a program that encourages them to participate in trying to improve their public services. If those who are willing to try find that they are able to act or convince others to act to improve their public services, they might find the experience helpful or empowering. If, however, those willing to try are unable to improve their public services or convince others to, they might find themselves discouraged or disempowered by the experience.

In 2015 and 2016, as part of a research project called Transparency for Development (T4D), several thousand people in 200 randomly selected rural communities across four regions of Indonesia and Tanzania were invited to attend a series of meetings in which a facilitator offered information about their maternal and newborn health care and encouraged them to plan ways that they might improve that care and then try that plan. In this paper, we use observations of program meetings and focus groups and interviews with those who participated during the program and two years after it to ask whether participants in these communities engaged in the discussions and deliberations in the meetings and in civic actions outside them, and whether they found the experience to be empowering and helpful for improving their maternal and newborn health care. (In other work (xx, xx, xx) we examine what those who were willing to participate were able to do to improve their maternal and newborn care, whether these efforts were sufficient to significantly improve measurable maternal and newborn health care outcomes, and what else the experiences were like for those who participated, other than an opportunity to improve their health care.)

The paper is organized as follows. Section 2 describes the program each community was offered. Section 3 discusses the mixed existing evidence for the two assumptions above about how people in these communities would react to and experience this program, and Section 4 describes five complementary perspectives we use to understand their participation in and experiences of the program: focus groups in all

[5] See Fox (2007b; 2015), Gaventa and McGee (2011), J-PAL (2011), Khagram, Fung, and de Renzio (2013), Lieberman, Posner, and Tsai (2014), and Kosack and Fung (2014).

two hundred communities with those who participated two years after the program ended, and, in a smaller group, qualitative observations of their participation in meetings and interviews with them and others in their communities before, during, and immediately after the program. Section 5 uses these observations and reflections to explore whether and how participants in Indonesia and Tanzania responded to the program, and Section 6 uses these perspectives to explore their perceptions of their civic efficacy. Section 7 concludes and offers several hypotheses for future work.

2. A Community-Led Scorecard Program to Improve Maternal and Newborn Health

The program offered in Indonesia and Tanzania was a community-led scorecard program designed to make the workings of each community's health care services more transparent and to encourage community members to deliberate on and then try a plan of activities that they thought would improve the quality of and access to those services. In each community, a facilitator from a local nongovernmental organization spent several weeks conducting a survey and gathering stories of experiences among community members with their maternal and newborn heath and health care. The facilitator then invited a group of interested participants who were broadly representative, particularly of non-elite community members, to attend six community meetings held over a period of approximately three-to-four months.

In the first two meetings, held back to back, the facilitator asked participants to discuss the results of the survey, statistics about their care relative to other communities, and stories of how other communities like theirs had improved their public services (See Appendix A for details on the information provided). The facilitator then asked those still interested to come back for a second meeting, held shortly after the first, in which they deliberated on a plan of activities that might improve those services or increase access to or use of them among the community. After these first two meetings, the facilitator and the participants presented this plan in an open meeting to anyone in their broader community who was interested in learning about it or becoming involved. The facilitator then departed the community. They returned for three more discussions, held approximately one, two, and three months later, at which they encouraged anyone still involved to meet and reflect on how their plan had worked, make adjustments, and plan for how to maintain their efforts once the facilitator left entirely.

The number, format, size, and content of the meetings in Tanzania and Indonesia was co-designed, piloted, and iteratively adapted over a two-year period by the research team and staff of nongovernmental organizations in Indonesia and Tanzania with experience in community health and transparency and accountability. The design was based on an earlier community scorecard program that had been shown to significantly improve maternal and newborn health on objective metrics (Björkman and Svensson 2009).[6] Like that program, the program in Tanzania and Indonesia was designed to focus specifically on improving

[6] In Tanzania, the Tanzania country office for the Clinton Health Access Initiative (CHAI), an organization focused on improving health service delivery in a number of domains; in Indonesia, PATTIRO, a research and policy advocacy organization focused on regional and local governance issues in a number of sectors (see Transparency for Development 2017; Arkedis et al. 2019).

maternal and newborn health care services, and to be maximally reliant on participants' own capacities by providing only information and discussion space, not outside resources or other incentives. Participants in Indonesia and Tanzania were provided with a few snacks at early meetings, and in Tanzania, small allowances were given to those who attended the first two of the six program meetings, but otherwise participants were offered no other resources, expertise, funding, authority, or other assistance.

In short, in 200 randomly selected communities across four regions of Indonesia and Tanzania, a trained facilitator offered information, time and space to meet, encouragement, and help making and adjusting plans. Facilitators were also trained to provide participants with nothing else: no guidance about what, if anything, participants should try, and no technical, relational, material, or financial resources to help them with whatever they decided to try.

3. Hypothesis and Alternatives

The existing evidence for this kind of approach to improving public services is mixed. As noted in the introduction, several studies suggest that transparency and accountability programs can create opportunities for citizens to learn about and act in ways that alleviate problems with their public services, leading them to be measurably more responsive and effective, while others find these programs to be insufficient, inappropriate, or even reinforce the problems they were designed to fix. Björkman and Svensson (2009), evaluating the community scorecard program in Uganda in 2004, produced extraordinary improvements in health and health care, including a 33% decline in infant mortality, despite providing rural villagers and clinic staff with only survey information and a facilitated forum for discussing and creating a plan for fixing problems the survey revealed. Banerjee et al. (2018) find substantial reductions in resource leakages in Indonesia from simply providing households with information about what they were supposed to receive from a subsidized rice program, increasing the subsides they received by an average of 26 percent. In India, a World Bank study (2018) finds that a program that provided information about health services and outcomes and facilitated community-wide meetings reduced the proportion of children who were stunted or underweight by 11 percent and increased immunization rates by 27 percent two years later. A recent evaluation of several hundred social accountability programs across 28 countries found that those that invited citizen engagement almost always improved service delivery in the area (Holland and Schatz 2016).

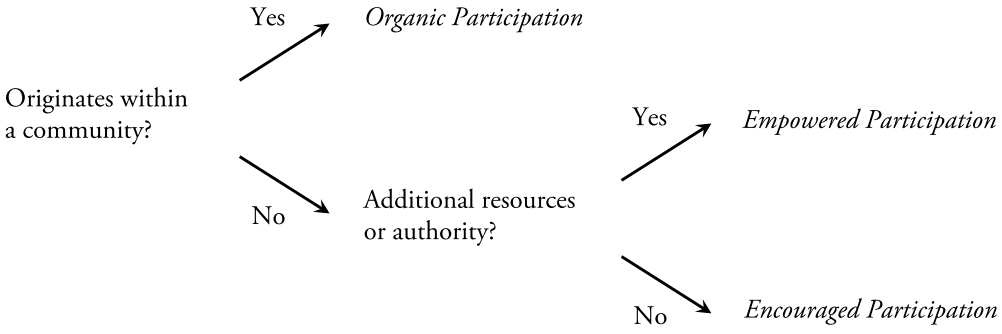
Yet other studies suggest the opposite. A randomized evaluation of an experimental citizen training and participation program in Karnataka, India in 2004, the same year as the program Björkman and Svensson (2009) evaluate, found no improvements in schools, health, sanitation, irrigation, roads, access to government programs, or tax compliance (Ananthpur, Malik, and Rao 2014): although all increased dramatically over the course of the program, they also increased dramatically in randomly selected communities who were not offered the program. Raffler, Posner, and Parkerson (2019) evaluate a large-scale replication of the Björkman and Svensson (2009), also in Uganda, and find only small effects on treatment quality and patient satisfaction and no effects on utilization rates or health outcomes, including

child mortality. Banerjee et al. (2010) find that a program that provided information to Village Education Committees in India on their role and activities and to community members on administering a basic reading test failed to increase the involvement of community members in education or the efforts of teachers, and had no effect on learning outcomes. Olken (2007) finds that a program that sought to encourage communities in Indonesia to monitor and hold officials accountable for resource leakage in road construction were unable to match local officials' ability to hide their malfeasance; only audits by higher level officials made a difference. A recent review of the experimental evidence in 21 countries from 25 evaluations of community driven development (CDD) programs—a related though distinct approach in which communities are also provided with resources with which to make improvements that they choose or in areas that they prioritize—finds that these programs very rarely improved public services, that those who come to the meetings rarely participate in discussions or decision-making, and that women and traditionally disempowered groups are particularly unlikely to participate (3ie 2018).

This mixed evidence base has led to a growing body of theoretical and empirical work examining the complex and contingent causal pathways between transparency, participation, and the responsiveness and effectiveness of public services.[7] *This paper focuses on two assumptions underlying the relationship: that in the communities where these programs are offered there are some people 1) who have the capacities to act or convince others to act in ways that could improve their community's public services, and 2) who would if they were encouraged to by a program offering only information and discussion space but not additional resources or authority.*

Figure 1 summarizes these distinctions.

Figure 1 – Source and Sufficiency of Participation to Improve Public Services



We begin with the first question in Figure 1: whether participation to improve public services can be encouraged by a transparency and accountability program. Civic participation is widely known to improve public services when it is organic, from and of a community, and those who are capable of this

[7] See, for example, Fox (2007b; 2015), Gaventa and McGee (2011), J-PAL (2011), Khagram, Fung, and de Renzio (2013), Lieberman, Posner, and Tsai (2014), or Kosack and Fung (2014).

kind of civic efficacy may also be unfamiliar with or have a tendency to ignore or discount that capacity.[8] But a program from outside the community may face numerous obstacles to encouraging participation to improve public services—particularly people in diverse, randomly selected communities who are encouraged to try to improve a specific public service like maternal and newborn health care.[9] Because access and quality of many public services have improved rapidly in recent decades and yet remain uneven across the globe, some communities have seen their services improve steadily even without their involvement, while for others, their benefits are hypothetical.[10] In either case, members of diverse communities may not perceive a given public service—even services like maternal and newborn health care that are generally believed to be so important to most people's lives that there is an international consensus on the general need to improve them—to be both sufficiently important and sufficiently underperforming that some are willing to use their scarce time and resources to try to improve that service.[11]

The second branch of Figure 1 describes a related though distinct question about these approaches: whether they can be empowering for improving public services even without additional resources or authority. The people who use public services might already have the capacity to act or convince others to act in ways that sustainably improve those services, even without additional resources or authority: organizing themselves to make improvements or working with staff, officials, other community organizations or leaders, journalists, nongovernmental organizations, or others able to contribute to improving their community's access to effective and responsive public services.[12] If so, even those skeptical that their efforts could improve their public services or that officials, providers, or others in their community would respond to their efforts might gradually discover latent capacities by trying and seeing changes they did not expect.[13] And programs that encourage them to make use of these capacities might

[8] Mansuri and Rao (2013), Baiocchi, Heller, and Silva (2011), Gaventa (2006). On ignoring or discounting more generally, see Khaneman and Tversky (1979) or Tversky and Khaneman (1991).
[9] Examining hundreds of participatory development programs Mansuri and Rao (2013) find that those in which participation was "induced" rather than organic were hardly ever effective, and Dupuy, Ron, and Prakash (2016) and Rahman, Miah, and Giessen (2017) note recent skepticism of nongovernmental organizations. The difference between organic and externally induced is in reality a dimension rather than a black and white distinction, and many scholars argue that participation that originates organically can still be aided by external resources, authority, or other assistance (e.g. Fox 2015; 2016; Keck and Sikkink 1998).
[10] See Kruk et al. (2016), UN General Assembly (2015), World Economic Forum (2015), Farmer et al. (2013), Hsia et al. (2012).
[11] See Pritchett (2015), Storeng and Béhague (2016) and Zulminarni et al. (2018) on priority differences between the international development community and the communities who receive development assistance.
[12] Hirschman (1970), World Development Report (2004), Fox (2007a; 2015), Joshi (2010), Gaventa and McGee (2011), J-PAL (2011), Kosack and Fung (2014).
[13] Agrawal (2005) describes an analogous process in changing attitudes and approaches toward environmental stewardship, in which "even if only a very small proportion of one's daily experience serves to undermine existing beliefs, over a relatively short period (such as a year or two) there may be ample opportunity to arrive at subject positions that are quite different from those held earlier."

avoid several problems of programs that offer resources, including dependency, elite capture, and incentives toward culturally or contextually inappropriate development.[14]

Yet even in communities where some are willing to try to improve their public services when encouraged to, those efforts may not be insufficient. The systems that deliver public services like maternal and newborn health care are complex and, aside from frontline facilities and providers, can be remote from those who use the services.[15] Transparency and accountability or social accountability programs in particular typically focus on some of the thorniest of problems with service delivery systems, many stemming from principal-agent failures, culture, and limited trust between citizens and the providers and officials delivering public services, that interfere with the implementation of policies and services that are responsive and effective to citizens' needs and preferences. These challenges might be insurmountable for those who are encouraged to try to improve their public services if they[16] or the state providing those services[17] are not also supplemented or supported with additional resources, authority, or other capacities. If so, those willing to try to improve their public services might find that they lacked the capacities, and the experience of trying, rather than confirming or revealing latent efficacy, might leave them less confident.

We examine these assumptions across 200 diverse communities in four regions of Indonesia and Tanzania who were offered the community-led scorecard program described in the previous section. It is important to note several important differences between this program and other contemporary transparency and accountability programs, differences that may limit the generalizability of our findings about the willingness of people to participate and the implications for their perceptions of efficacy to other approaches to transparency and accountability or social accountability.[18] Among these differences are the "community," which in the program in Indonesia and Tanzania was a lived rural community such as a village or small town, not the regional, national, or international communities engaged in other transparency programs or initiatives.[19] Second, the program was designed to avoid presuming the particular problems each community might be facing with their maternal and newborn health care by offering information and discussion opportunities of relevance to a range of ways of improving the quality of care available or access to it and awareness or knowledge of it among the community, rather than information specific to a particular problem with that service, such as absenteeism rates, drug stockouts, or evidence of

[14] For example, Mansuri and Rao (2013), Moyo (2009), Easterly (2007; 2006), Scott (1999), Ferguson (1994). Earlier works on the dependency that resources from foreign assistance can encourage include, for example, Cardoso and Faletto (1979), Bauer (1972), Prebisch (1950), and Singer (1950).
[15] See Lipsky (1980) and Zacka (2017), among many others.
[16] See Baiocchi, Heller, and Silva (2011), Gaventa (2006), Fung (2006), Fung and Wright (2003).
[17] See Andrews, Pritchett, and Woolcock (2017), Fox (2016; 2015; 2007b; 2005), Bessley and Persson (2010), Grindle (1997; 2007). In the wake of earlier calls for international philanthropic efforts to improve public services directly (e.g. Sachs 2005), many public services are also increasingly reliant on non-state support.
[18] See discussions in Fox (2015), Kosack and Fung (2014), Joshi and Houtzager (2012), Gaventa and McGee (2011), J-PAL (2011).
[19] For example, the International Budget Partnership's Open Budgets Survey or Transparency International's Corruption Perceptions Index. (See, e.g., Hahn, Reimsbach, and Schiemann 2015; Khagram, Fung, and Renzio 2013; Rose-Ackerman 1999). In fact, to minimize "spillover" effects that would interfere with the experimental design, the programs in Indonesia and Tanzania were offered in rural communities that were typically far apart.

malfeasance.[20] Third, the program was designed to avoid presuming the political, economic, or social context, by not suggesting or incentivizing any particular approach or kind of civic action, such as petitioning officials for reforms or budgetary reallocations, working with frontline providers, submitting a complaint, confronting public officials with evidence of malfeasance, or exiting to a private provider,[21] and instead offering a range of stories of how nearby communities had improved their public services and emphasizing that the participants should use their understanding of their context to create a plan they thought would be effective. Finally, facilitators in the program in Tanzania and Indonesia did not only provide information on its own, without encouraging participants to do something to try to fix any problems the information suggested, or pay participants, with the exception of small allowances provided to participants in the first two program meetings in Tanzania.

Yet the program in Indonesia and Tanzania does offer an opportunity to ask whether a transparency and accountability program can encourage civic participation in diverse communities to improve a public service by offering only information and opportunities for facilitated discussion and planning of civic engagement, not additional resources or authority, as well as whether the experience is discouraging and disappointing or empowering and helpful for improving that service. Facilitators in Indonesia and Tanzania were employed by a local nongovernmental organization and some were from neighboring communities, but they were not typically from the community, and participants were generally aware that the program itself was partly supported by external organizations. The information and space for discussion and deliberation that facilitators provided might have been sufficient to encourage interested community members to engage in civic participation through which they used latent capacities to improve a valued public service. But in many places people may not have been willing and able to take time from work and family to try to understand problems with a particular public service and then to try to improve that service. It might also have been unreasonable to expect those who were willing to try to find themselves able to improve that service. Many might have instead ignored the program, and those willing to engage might have found the experience unhelpful or discouraging, leaving them less confident in their civic efficacy than before they tried.

4. Methods, Data, and Contexts

[20] For example, many transparency and accountability or social accountability programs include an informational component on a specific problem that not all communities face, such as nurse or teacher absenteeism (absenteeism studies), leakages along the way from the treasury to the school or clinic (Public Expenditure Tracking Studies), malfeasance or corruption (social audits), or a lack of time and space for providers and community members to work out differences and engage in joint problem-solving (Community Score Cards).

[21] Many transparency and accountability programs that encourage civic participation also prescribe a particular action or set of actions for participants to engage in to alleviate problems with public services: for example, collaborative problem solving, as in programs that include an "interface" meeting between community members and providers following the discussion of the information (Björkman and Svensson 2009), or naming and shaming, as in the case of programs that include a "social audit," in which community members have the opportunity to confront public officials with evidence that they have not done what government documents claimed was done (Jenkins and Goetz 1999).

The remainder of the paper relies on five varied perspectives on whether participants in the communities who were offered the program largely ignored it or engaged meaningfully in the meetings and civic participation, and whether or not those who participated seemed to find their participation helpful for enabling them to improve their public services:

1. In 81 communities of the 200 where the program was offered—41 in Indonesia and 40 in Tanzania—trained observers attended three of the meetings and answered a series of questions about engagement, discussion, and decision-making in those meetings: the first and second meetings, at which participants discussed the information provided and deliberated on activities to pursue, and the third and final follow-up meeting, at which participants discussed their progress over the previous three months and made plans to sustain their activities after the facilitator had left. These meeting observations provide data on participation in meeting discussions, suggestive evidence from those discussions of whether participants engaged outside the meetings in any of the activities they had planned, and suggestive evidence of their optimism or skepticism in those discussions whether from their perspective their efforts could and were improving their community's maternal and newborn health care.
2. In 65 of these communities—41 in Indonesia and 24 in Tanzania—interviewers asked follow up questions of several participants as well as of those with whom they engaged as part of their activities: providers, officials, their neighbors, and others. These interviews help to verify that the activities participants described in meetings actually occurred, and provide an indication of challenges participants may have faced, their and others' motivations in participating in any efforts to try to improve their health care, and their perceptions of whether they had improved their health care. Interviews in Indonesia were conducted after the program ended in all 41 communities where observers attended meetings; in Tanzania, interviews were conducted in 24 of the 40 communities where observers attended meetings after the program as well as two months later.
3. In 35 of these communities—16 in Indonesia and 19 in Tanzania—a second observer noted the frequency with which meeting participants made distinct contributions during the discussions and activities that each discussed engaging in outside the meetings.[22]
4. In these 35 communities, the two observers also interviewed all participants before the first meeting and after the last about their perceptions of individual perceptions of their capacities to improve their communities. To allow more valid interpersonal comparisons among responses, interviewers also asked participants to compare their own civic

[22] We thank Jane Mansbridge for suggesting this method of understanding individuals' participation.

efficacy to a set of three anchoring vignettes in which a person in a community like theirs tries to improve their school (Masset 2015; King et al. 2004).[23]

5. Finally, two years after the program ended, an interviewer invited participants in all 200 communities to a focus group in which they were asked to reflect on the program and any activities they remembered, including challenges they had faced and any changes they had seen as a result. They also asked whether participants thought that their efforts had improved their health care overall, whether they were still trying to improve their care or their communities in other ways, any personal benefits or costs from participating, and overall whether or not they were glad that they had participated.

These data sources are designed to follow the recommendations of recent scholarship in comparative methodology by exploring participants' engagement with and experiences of this program with data collected at varied times and from distinct yet overlapping perspectives, such that each partly compensates for the disadvantages and biases of the others in reliably reflecting how participants responded to the program.[24] In particular, meeting observations focused on general meeting dynamics: participants' interest and engagement in the conversations, the role of the facilitator, participants' discussions and deliberations at the start of the program around planning activities, and participants' reflections at the end of the program on activities they had taken, challenges they had faced, and plans for the sustainability of their activities. In early meetings and the last program meeting, observers were also asked for their perceptions of how skeptical or optimistic participants appeared to be that their efforts would make a difference; these observations complement participants' perceptions of their civic efficacy as reported in interviews and focus groups: the former entail observer bias but are also less susceptible to other biases inherent in asking participants themselves about their experiences and their perceptions of their capabilities. We also complement these with participant's recollections and reflections two years after the program: the latter involve recall bias but are less susceptible to biases or incentives toward socially desirable responses than during the program. The focus of meeting observers on the general dynamics of each meeting also meant that they focused less on individuals' participation in the meetings, and differences in participation among education, gender, and age groups often associated with empowerment and marginalization. We explore these instead with interviews and with systematic observations of their participation in meetings, including the number of times each person spoke distinctly from the rest of the group or described volunteering to do, being assigned, or having tried activities outside of the meetings. To increase representativeness of all data sources, the communities where meetings were observed and participants

[23] We thank Anuradha Joshi for suggesting this approach to measuring empowerment. There is an analogous concept in psychology of individuals called "self-efficacy": an individual's belief in their capabilities (Bandura 1977; 1982). It is generally assessed similarly: by asking individuals how capable they think they are of achieving a goal or task.
[24] This general approach draws from scholarship in comparative methodology, including Mill (1843), King, Keohane, and Verba (1994), Lieberman (2005), Seawright (2016) among others, as well as in international development, including scholarship at the World Bank (e.g. Woolcock 2013; Ananthpur, Malik, and Rao 2014) and the UK Department for International Development (e.g. Stern et al. 2012). Interview and observation protocols are available at t4d.ash.harvard.edu.

interviewed were selected randomly from the same national and regional stratifications as those who were offered the program as part of the broader randomized controlled trial.

As contexts in which to ask how people engage with and experience a transparency and accountability program,[25] Tanzania and Indonesia differ on many economic, political, cultural, and geographic dimensions, as well as in the reach and capacities of their health care systems. Indonesia is a democracy that has experienced multiple peaceful transfers of power since the fall of an authoritarian regime in 1998 (Freedom House, 2018); in Tanzania, by contrast, multiple parties regularly contest elections but one party has retained power for over half a century (Freedom House, 2018). By the World Bank's classifications, Tanzania is a low-income country; Indonesia is a lower-middle-income country, with four times as many economic resources per capita and commensurately higher living standards (World Bank, 2017).[26] These resource differences are reflected in the health care systems in the two countries. The program was offered to communities in two provinces on two of Indonesia's eighteen thousand islands, Banten and South Sulawesi, and in two regions of Tanzania, Dodoma and Tanga. In Tanzania, the primary public facility is a dispensary, while the Indonesian system involves a range of providers: each sub-district has a large public health center offering comprehensive public health services, including delivery services (puskesmas), and which overseas a network of smaller facilities and a community-based program staffed by a midwife and local volunteers that offers monthly pre and antenatal services to pregnant women and health monitoring and vaccination programs for children younger than five. In Indonesia, 99% of communities offered the program had a smaller health center overseen by a facility (puskesmas), a village midwife operating under the supervision of the puskesmas, a birthing clinic, or a private provider located within the village. In a survey prior to the program of puskesmas whose catchment areas included communities offered the program, each had close to 57 staff, nearly 14 in the maternal/delivery unit alone, nearly all had electricity, and 96% used water from an improved source. In Tanzania, 62% of communities who were offered the program had a dispensary, the primary public facility, in the community, and the dispensaries whose catchment areas included communities offered the program were on average both smaller than those in Indonesia—6-7 staff, on average—and far less likely to have electricity (22%) or regular running water (44%). In a surveys of recently pregnant women in each community in Indonesia before the program, 82% said they had more than one provider or facility from which they could seek care; in Tanzania, 74% of recently pregnant women said they only had one health facility to choose from.[27] Nearly all in both countries had received some antenatal care, but in Indonesia, 90% had had the recommended four antenatal care visits; in Tanzania, 43% had.

Yet despite their diversity, the two countries both share important characteristics of places where transparency and accountability programs are often offered, including wealth alongside acute poverty,

[25] See King, Keohane, and Verba (1994); Mill (1843).
[26] In a household survey in communities before the program began, 13% in Tanzania had access to electricity compared to 99% in Indonesia. Most houses in Indonesia were of stone (48%) or wood (37%); in Tanzania, 71% lived in dwellings of mud.
[27] See Arkedis et al. (2019) and Transparency for Development (2016) for more detailed discussions of similarities and differences between the health systems in Banten, South Sulawesi, Dodoma, and Tanga.

ethnically heterogeneous citizenries, politics with substantial patron-clientalism, and recent experiences of colonialism and undemocratic governance. Although average health care quality and access differed between the two countries, they also shared decentralized health systems that expanded and improved dramatically in recent decades but whose quality and availability remained uneven and limited in many places.[28] In surveys of recently pregnant women before the program began, just 22% in Indonesia and 31% in Tanzania described receiving three standard components of antenatal care: having their blood pressure checked, having a urine sample drawn and receiving a report of the results, and having a blood sample drawn and receiving a report of the results. 45% in Indonesia and 44% in Tanzania said that their most recent birth was not at a health facility. During the period of the program, both countries were governed by presidents who placed improving public services, particularly health care, near the top of their agendas.[29]

5. Participation In and Outside Meetings

Observations of meetings and interviews and focus groups with those who participated suggest that in a large majority of communities, engagement in meetings was broad, deep, and sustained; that participants frequently though not always made decisions themselves rather than relying on the facilitator; and that some who participated in the meetings also tried some of the approach they had planned to improve their maternal and newborn health.

Participants had been trained to invite about 15 people to the meetings they held. Although facilitators were from outside the communities and offered those who attended the meetings no additional resources, authority, or other incentives, an average of 14 people attended the three meetings that researchers observed (max: 28; min: 3). In most communities, several others attended as well: members of other volunteer organizations, village leaders, spouses, police, and others who were curious. Facilitators in both countries described disappointment among many participants when they first learned that they would receive no payment in exchange for their participation,[30] and in most there was substantial early attrition as less interested participants dropped out. Yet in all communities a group generally emerged, either from among the original participants or from new volunteers, who participated throughout: in Indonesia an

[28] Yahya and Mohamed (2018), WHO (2017). Over 90% of Tanzanians live within 5 km of a public health facility, according to the government (Ministry of Health and Social Welfare 2008). 93% of the Indonesians report easy access to a primary care facility (either a sub-district health center or private practitioner) in 2011, and 90% have easy access to a delivery facility (Vothknecht and Sparrow).

[29] In Indonesia, Joko Widodo; in Tanzania, John Magufuli. As we note later in this paper and in Arkedis et al. (2019), in the four regions in which the program we consider here was offered in the years during and after it, the quality and use of maternal and newborn health care services continued to improve substantially.

[30] This lack of "per diems" was noted repeatedly. Even after the program had concluded, participants brought the issue up in 32 of the 41 communities in Indonesia and 17 of the 24 communities in Tanzania in which interviewers spoke with them. We observed it from other perspectives as well: in addition to the enumerators, interviewers, and observers already discussed in the paper, scholars lived among a number of communities before, during, and after they were offered the program and developed intensive ethnographic studies of how those who participated reacted to and made sense of the program and their participation in it; these studies suggest that expectations to be paid per diems were common and important to how participants reacted to and understood the program in nearly all communities, and may have been decisive in some (see [book]).

average of 14 attended the first program meeting, 12 attended the second and 11 attended the last; in Tanzania, an average of 15 people attended the first and second meetings and 10 attended the last.

Observations of meeting discussions and of individuals' participation in those discussions also both suggest that those attending were broadly engaged. In the 81 communities where observers attended meetings, they noted whether those attending appeared to be engaged in general, across the whole meeting, as well as at several specific points during the discussions, and whether they brought up any local stories or examples from their community of the more general topics of conversation. Table 1 summarizes engagement in meetings; see Tables B1-B3 in Appendix B for details and definitions. Based on these observations, participants in the first meeting appeared to be completely or mostly unengaged in just seven of the 81 communities (9%): two in Indonesia and five in Tanzania. In a majority of communities in both countries, most specific points in discussions were engaged and involved local examples (see Table B2 in Appendix B).[31]

Broad and sustained engagement is also reflected in observations of individuals' participation (Table 1 and Table B3 in Appendix B). These suggest that very few of those who attended the meetings sat quietly. In Indonesia, 89% of participants in the first meeting and 94% of participants in the final meeting made some distinctive contribution to the discussion; 57% of participants in the first meeting and 71% of participants in the final meeting made five or more. In Tanzania, 91% of participants in the first meeting spoke distinctly, and 55% made five or more distinct contributions. By the final meeting, participation in Tanzania had declined somewhat, one indication of an increase in skepticism among a substantial number of participants in Tanzania that we see repeatedly below from other perspectives. Yet 68% of participants in the final program meeting still made some distinctive contribution and 38% spoke distinctly five or more times. Observations of individuals' participation also suggest that in both countries, and particularly in Indonesia, most of those who spoke in discussions were women.[32] Those who participated were also from a wide cross-section of age and educational groups. In Tanzania, most of those who spoke had at most a junior high school education and tended to be relatively older than in Indonesia: about a quarter of those who spoke were elderly. In Indonesia, the average level of education was higher, yet less than half of those who spoke in discussions had senior secondary or tertiary education, and very few were elderly.

In both countries, these observations of individuals' participation also suggest that a core group was particularly active in the meetings in most communities.[33] In Tanzania, observations of individuals' participation suggest that the average participant spoke distinctly at least seven times; the most talkative, 40 times in the first meeting and 46 times in the last. In Indonesia, the average participant spoke 16 times in

[31] One notable exception was the discussions in the first meeting of stories of other communities improving their public services: observers thought participants were engaged in these discussions in only nine communities in Tanzania and only two in Indonesia.

[32] In this way participation in the programs in Indonesia and Tanzania was unlike studies of community driven development programs reviewed recently by (3ie 2018): their review finds that women and those from traditionally disempowered groups are particularly unlikely to speak.

[33] Mansbridge (1980) notes a similar dynamic in community town hall meetings in the northeastern U.S.

the first meeting and 32 times in the last; the most talkative, 114 times in the first meeting and 186 in the last. Yet observations of meeting discussions suggest that this disproportionately engaged core group seldom seemed to control or otherwise dominate the conversation to the exclusion of others at the meeting. Observers watched for this sort of small group dominance at six specific moments in the meetings: when participants were deciding problems to focus on, hearing stories of nearby communities like theirs who had improved their public services, discussing in small groups and then altogether what they could do to improve their care, creating a plan of activities they could try to improve their care, reviewing their progress since the last meeting, and planning for the sustainability of their activities once the facilitator left for the final time. Observers noted a small group dominating a majority of these discussions in the first program meeting in only five communities in Indonesia and three in Tanzania, and in the last meeting in only nine communities in Indonesia and two in Tanzania (Table 1). Indeed in the majority of communities in both countries, observers did not think a small group dominated any of the discussions in the three meetings they attended (Table B1 in Appendix B).

The meetings were designed to offer opportunities for participants to decide for themselves what information was useful, what aspects of their care to focus on, and what to try to improve those aspects. But their discussions or deliberations might also have been directed by facilitators toward issues or decisions that they, not the participants, thought would be most effective or otherwise advisable.[34] Likewise participants might have relied on the facilitators as sources of direction or knowledge, rather than on their own experiences and intuitions of what would work in their communities. Observers watched for control by or reliance on the facilitator during all the same moments in the discussions when they watched for dominance of a small group of participants, as well as when the facilitator was offering them information they had gathered on the state of maternal and newborn health care in the community. These observations suggest that more often than not, participants conducted discussions and made decisions themselves rather than relying on the facilitator (Table 1). There were moments of both reliance on and control by the facilitator in some communities, but in almost all, participants decided for themselves the problems to focus on and approaches to take in resolving them. Observers noted participants relying on the facilitator in all discussions in the first meeting in only three communities in Indonesia and one in Tanzania, and in the last meeting in only nine communities in Indonesia and none in Tanzania.

[34] These patterns of participation in decision-making are also distinctive from studies of community driven development programs in the aforementioned systematic review (3ie 2018), which finds that those who come to the meetings, and particularly women and traditionally disempowered groups, participate disproportionately rarely in decision-making.

Table 1 – Engagement in meetings

	Indonesia (41 communities)			*Tanzania (40 communities)*		
Meeting:	*1*	*2*	*6*	*1*	*2*	*6*
Proportion (%) of communities in which						
participants were somewhat or very engaged	95%	93%	88%	88%	93%	80%
most discussions* were completely or extensively dominated by a small group of participants	12%	17%	22%	7%	0%	4%
in most discussions*, participants relied completely or extensively on the facilitator in making decisions	7%	17%	25%	4%	0%	0%
in most discussions*, decision-making was completely or extensively controlled by the facilitator	3%	22%	22%	2%	0%	0%
Proportion of those attending who made						
at least one distinctive contribution to the discussion	89%	95%	94%	91%	89%	68%
five or more distinctive contributions	57%	70%	71%	55%	50%	38%
Of those who spoke distinctly						
Female	74%	74%	74%	55%	54%	51%
Junior Secondary education	17%	20%	18%	74%	73%	71%
Senior Sec or Tertiary education	49%	44%	48%	12%	12%	15%
Age 20-55	90%	89%	87%	70%	73%	62%
Age >55	3%	2%	3%	31%	27%	37%
Times average participant spoke	16	21	32	7	7	7
Times most verbose participant spoke	114	202	186	40	48	46

Notes: For definitions and more detailed statistics on engagement in meetings and small group and facilitator involvement, see Table 1 in Appendix A. Table 2 in Appendix A details engagement in specific moments in the discussions and deliberations across the 3 meetings. For definitions and more detailed statistics on individuals' participation in meetings, see Table 3 in Appendix A. * Observers were asked about small group dominance and the involvement of the facilitator during six moments in the discussions: when participants were deciding problems to focus on, hearing stories of nearby communities like theirs who had improved their public services, discussing what they could do to improve their care, creating their plan, reviewing their progress since the last meeting, and planning for the sustainability of their activities once the facilitator left for the final time.

Finally, meeting observations, interviews with participants and those with whom they tried to engage, and focus groups with participants two years after the program offer clear evidence that in almost all communities, some of those who attended meetings also tried some of the activities that they had planned to try to improve their maternal and newborn care. In every community in Indonesia and in all but one community in Tanzania, those who attended the final program meeting engaged in detailed discussions about their attempts prior to that meeting to try at least some of what they had planned. In all we identified

43 distinct approaches.[35] Some described trying to meet with providers to work out difficulties; others put up suggestion or complaint boxes; others made requests of their governments regarding the availability of drugs and supplies, the availability of transportation to the facility, the quality of the road leading there, and other aspects of care that needed improvement. Some asked for more skilled health providers to be assigned to their community, others for entirely new health facilities; participants in a few went so far as to acquire land and start building. Many tried to educate their neighbors about the importance of giving birth and seeking ante and post natal care in a modern health facility; some tried to adopt rules or pass laws about use of facilities; some organized and raised funds for transportation pools to help those who needed to get to the facility in a pinch. In all but one of the 65 communities in which interviewers spoke subsequently with those with whom participants had planned to engage, these individuals independently confirmed that participants had tried at least some of these activities; in all, interviewers were able to verify 85% of the activities that participants had described trying by talking to those whom participants had planned to engage (84% in Indonesia; 86% in Tanzania). Even two years after the program had ended, participants in every community in Tanzania and 97 of the 100 communities in Indonesia remembered specific activities sufficiently to allow them to engage in lengthy conversations reflecting on which had been more or less successful and why.

These responses suggest that a transparency and accountability program can encourage meaningful deliberation and participation in improving a public service. Facilitators were not from these communities and, with the exception of allowances offered in Tanzania for participation in the first two program meetings, they offered no resources, authority, or other incentives to those who attended the meetings either to participate in them or to augment their existing capacities as members of their communities and citizens of Indonesia or Tanzania. Yet observations of meeting discussions and of individual participation in meetings, interviews with participants and those with whom they tried to engage, and focus groups with participants two years after the program all suggest that in both countries, participants in most communities took substantial and sustained advantage of the opportunities the program offered to discuss and deliberate on how they could improve their maternal and newborn health care, and that they almost always tried at least a part of the approach they had planned.

6. Perceptions of Civic Efficacy

Participants who were willing to try to improve their maternal and newborn health care services might have begun the program thinking already they would be capable of making improvements and had that expectation confirmed as they tried. The experience may also have changed their perceptions of their efficacy. Some may have started skeptical but become more confident in the course of trying that they that capability, so that by the program's end they were optimistic that their efforts were making a difference. On the other hand, those who tried might have found themselves lacking that capacity and ended the program

[35] For more on the approaches participants planned to improve their care, see [SAP report].

skeptical that their efforts were making any difference and less confident in their efficacy than when they started. This final section offers evidence from interviews, focus groups, and meeting observations before, during, immediately after the program, and two years later, which all suggests that participants in Indonesia and Tanzania began with varied perceptions of their civic efficacy, and that over the course of the program, about a third became skeptical and remembered participating as costly or a waste of time. Yet many more were or became confident that their efforts were improving their health care and in their civic efficacy more generally. And eventually, most described having improved their health care, having benefiting from participating, and as or more confident in their civic efficacy than before the program began.

Before the Program

Participants began the program with varying confidence in their civic efficacy. Prior to the first program meeting, interviewers in 35 communities asked each participant about their perceptions of their civic efficacy, including this question:

> "I would now like you to think of improvements of any kind that you would like to make to improve life in your village, for yourself and others—for example, improving garbage collection to keep the village clean, fixing a bad road, organizing a watch group to keep the neighborhood safe, or anything else that you think would improve life in this village for yourself and others. Do you feel that you have the power to help make these kinds of improvements to life in this village, for yourself and others?"

Each was then asked how able they perceived themselves to be on a 4-point scale from "completely able" to "unable."

Responses in these interviews suggest an important difference prior to the program in the confidence of participants in Indonesia and Tanzania. In Indonesia, participants began substantially less confident: two-thirds of respondents responded to this question with a 1 or 2 (somewhat or completely unable) on the 4-point scale (Table 4). In Tanzania, most participants began highly confident: only 3% responded with a 1 or a 2.

Yet this difference faded once the program began. In both countries, participants appeared confident that they would be able to improve their communities' maternal and newborn health care. At several moments in the discussions of first two meetings, observers were asked to note whether participants seemed skeptical that:

1. nearby communities like theirs had been able to improve their public services;
2. they would be able on their own to plan and undertake activities that would improve their maternal and child health care; or
3. the activities they planned would help alleviate problems with their care.

Table 2 shows the proportion of communities in which observers noted either complete skepticism (all or nearly everyone was skeptical) or widespread skepticism (more than half were skeptical) among meeting participants. These observations reflect widespread optimism in most communities in both

countries at the start of the program. In only four communities in Indonesia and two in Tanzania, all or a majority of those attending meetings seemed skeptical at hearing of other communities improving their public services. In just six communities in Indonesia and five in Tanzania, all or a majority seemed skeptical that they would be able to improve their maternal and newborn care in the first meeting. And by the time they had created their plans in the second meeting, participants remained skeptical that their plan would allow them to improve their maternal and newborn care in only two communities in Tanzania and no communities in Indonesia.

Table 2 – Skepticism

Proportion of communities in which:	*Indonesia*	*Tanzania*
Meeting 1: Information		
All or a majority of participants seemed skeptical at the stories of other communities improving their public services	10%	5%
All or a majority seemed skeptical that it is possible for them to develop and realize ways to make improvements without outside assistance	15%	13%
Meeting 2: Planning		
All of a majority of participants seemed skeptical that the approach they had developed in small groups would help to alleviate the problems they had chosen to focus on	2%	5%
All or a majority seemed skeptical that the final plan of activities would help alleviate the problems they had chosen to focus on	0%	5%
Meeting 6: 3rd Follow Up		
All or a majority of participants seemed skeptical that their efforts would sustain improvements	12%	53%
Everyone seemed skeptical that their efforts would sustain improvements	0%	43%

Over the Course of the Program

To what extent did this optimism endure over the course of the program? During the third and final follow-up meeting, after which the facilitator would leave the community for the last time, observers were asked to note whether all or a majority of those attending were either skeptical or optimistic that they could sustain any progress they had made. Table 3 summarizes changes in optimism and skepticism over the course of the program.

In Indonesia, participants in 31 of the 41 in which observers attended meetings started out generally optimistic that they would be able to improve their maternal and newborn health care and remained optimistic at the end that they would be able to sustain the improvements in that care. In one other, participants started out skeptical and remained so at the final meeting. In the other nine communities, participants changed their perceptions. In four, most or all participants started optimistic but had become skeptical after three months of attempting to improve their care. In five others where all or most

participants had started the program appearing skeptical, they appeared to meeting observers to have become optimistic by the end, suggesting that they had found themselves to have latent capacity they were skeptical of at the start.

In Tanzania, participants in most communities also started out optimistic, but observations of the last meeting suggest that after attempting their approach for several months, many more than in Indonesia had become skeptical (Table 3). In 17, participants both started out and remained optimistic, and in three others participants started out skeptical and remained skeptical at the end of the program. In 20 other communities, participants changed their perceptions. In the other two of the five communities where participants had started out skeptical, they seemed to have become optimistic that they would be able to sustain improvements in their care. But in 18 where participants had started out optimistic, they had become skeptical by the end of the program, suggesting that rather than confirming or revealing latent efficacy the experience had left them less confident that they could improve their care than when the started. In all, at the end of the program participants in a slight majority of communities in Tanzania—21 of the 40 in which observers attended meetings—seemed skeptical that they would be able to sustain improvements in their care. In 17 of these, not a single participant seemed optimistic (Table 2).

Table 3 – Changes in Skepticism and Optimism in Program Meetings

Start of program	*Conclusion of program*	
	Skeptical	*Optimistic*
Optimistic	*Disappointment* Indonesia: 11% (4) Tanzania: 51% (18)	Indonesia: 90% (31) Tanzania: 49% (17)
	Indonesia: 85% (35) Tanzania: 88% (35)	
Skeptical	Indonesia: 17% (1) Tanzania: 60% (3)	*Discovery* Indonesia: 83% (5) Tanzania: 40% (2)
	Indonesia: 15% (6) Tanzania: 13% (5)	

Interviews with individual participants show similar differences between the two countries, and also suggest that in Indonesia, the experience led many individual participants to increase their perceptions of their general civic efficacy, while the more mixed experiences in Tanzania led far fewer participants to change their perceptions. After the last program meeting, in the same 35 communities where they asked participants about their perceptions of their own civic efficacy prior to the first program meeting, they again asked participants how able they perceived themselves to be on the same 4-point scale, from "completely able" to "unable" (Table 4).

Table 4 - Civic efficacy in individual interviews

	First program meeting				*Last program meeting*						
	Completely able ⇔ Unable				*Completely able ⇔ Unable*						
	4	3	2	1	4	3	2	1	*Increase*	*Unchanged*	*Decrease*
Attended both meetings											
	46%	23%	24%	6%	52%	26%	20%	2%	29%	57%	14%
Indonesia	9%	30%	49%	12%	28%	37%	34%	1%	46%	46%	7%
Tanzania	80%	18%	2%	1%	74%	16%	5%	2%	13%	67%	20%
Attended one meeting only											
	42%	22%	27%	9%	41%	25%	32%	3%	-	-	-
Indonesia	6%	26%	50%	18%	16%	30%	50%	5%	-	-	-
Tanzania	81%	18%	2%	0	75%	19%	6%	0	-	-	-

473 respondents in meeting 1; 306 in meeting 6. 230 attended both meetings; 243 attended meeting 1 only; 76 attended meeting 3 only.

These interviews suggest that in Indonesia, where most participants had begun the program perceiving themselves somewhat or completely unable to improve their communities, nearly half (46%) of participants told interviewers that they were more confident in their civic efficacy at the end of the program. Prior to the program, 60% of respondents responded with a 1 or 2 on the 4 point scale; after the final program meeting, only 35% responded with a 1 or 2.

Like the observations of meetings, these interviews suggest that over the course of the program, experiences in Tanzania were more mixed. Before the program, very few participants responded with a 1 or 2 on the 4-point scale and 80% responded that they were "completely able" to improve their communities. Among the 20% of participants in Tanzania who had started out less than fully confident, 13% responded after the program that they were more confident in their efficacy. But 20% described being slightly less confident in their general civic efficacy—although most remained confident: only 7% of participants responded after the program that they were somewhat or completely unable to improve their communities.

Finally, these interviews suggest that those in Tanzania whose confidence had increased over the program were those who were more engaged in the program, while in Indonesia, confidence increased among both those who were relatively more and less engaged. In Table 5, we model self-reported efficacy (*civic efficacy*) of those who participated all the way through the program, attending both the first meeting and the last, as a function of their time participating in the program and the intensity of their participation, as well as their age, education, gender, and where they live, as follows:

$$civic\ efficacy_i = \alpha + \beta 1 time_i + \mathbf{\beta_2 Z_i} + \beta_3 P_i + \beta_4 (time_i \times P_i) + \eta_i + \varepsilon_i \quad (1)$$

where *time* is a dummy variable indicating that the interview was conducted after the final meeting rather than before the first; **Z** is a vector of three characteristics that may also be important to an individual's perceptions of their civic efficacy: age (in quartiles), gender, and level of education (in quartiles); *P* is an

index that proxies for the intensity of individual *i*'s participation in the first or last program meeting[36]; *time* x *P* is an interaction term to estimate whether any association between *civic efficacy* and *time* depends on how engaged participants were in meetings and in civic activities outside them to try to improve their care; and η are fixed effects for community to proxy for unobserved contextual factors that might be associated with the perceptions of civic efficacy of those who live there. *i* indexes for participant; ε is a random error term.

Table 5 – Civic efficacy: ordered probit estimates of equation (1)

	Both countries		*Indonesia*		*Tanzania*	
End of the program (*time*)	0.425***	0.0486	0.990***	1.030***	-0.182	-1.164***
	(0.119)	(0.217)	(0.164)	(0.295)	(0.187)	(0.388)
Participation Index (*P*)		0.0204		0.136**		-0.264**
		(0.0551)		(0.0681)		(0.109)
time x *P*		0.116*		-0.0487		0.383***
		(0.0651)		(0.0791)		(0.133)
Controls						
Age (thirds)	-0.0193	-0.0557	0.0734	-0.00609	-0.0711	-0.0527
	(0.107)	(0.109)	(0.159)	(0.163)	(0.164)	(0.168)
Education (quartiles)	0.333***	0.311***	0.403***	0.366***	0.37	0.488*
	(0.0896)	(0.0909)	(0.100)	(0.101)	(0.256)	(0.268)
Gender (female)	-0.354**	-0.327**	-0.286	-0.263	-0.464**	-0.451*
	(0.150)	(0.151)	(0.208)	(0.208)	(0.235)	(0.243)
Fixed effects for community	Y	Y	Y	Y	Y	Y

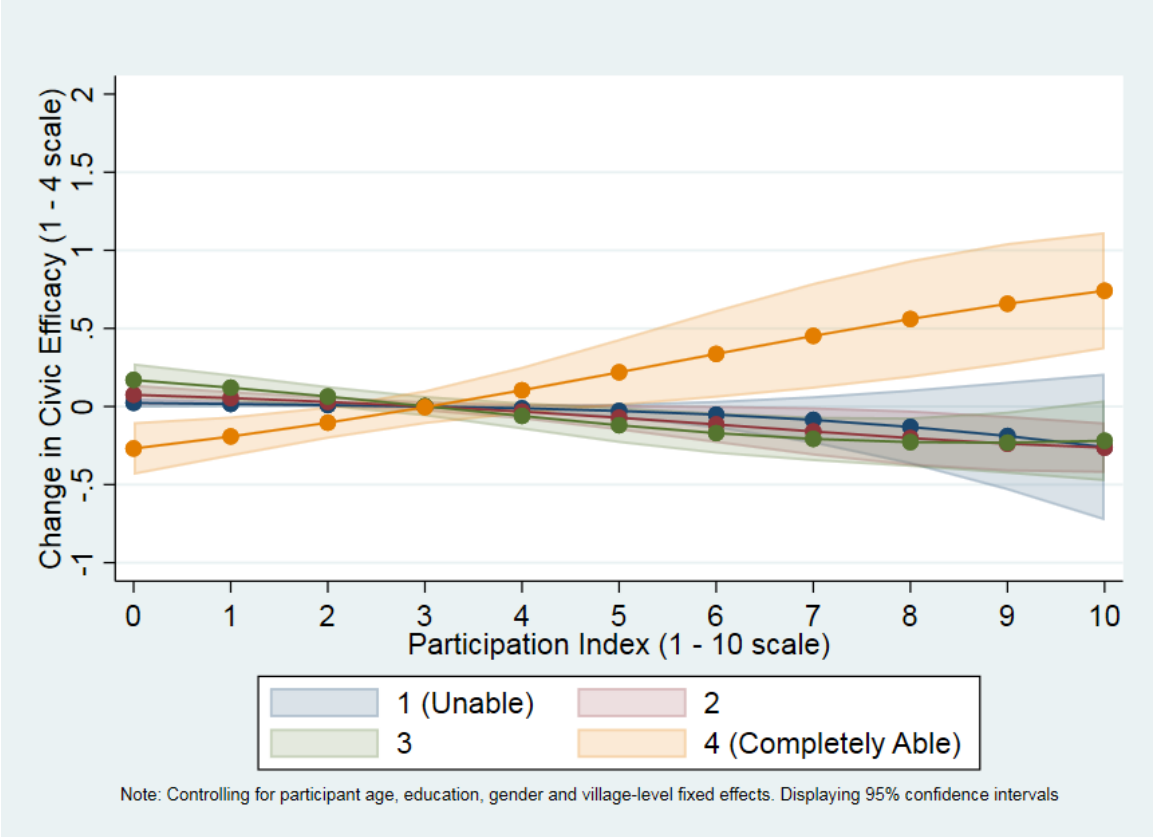
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Sample restricted to respondents in both meetings.
Both Countries: Respondents: 219. Communities: 31.
Indonesia: Respondents: 108. Communities: 15.
Tanzania. Respondents: 111. Communities: 16.
See Appendix C for estimates with alternative indices for participation.

Overall, the period of the program is strongly associated with increased perceptions of civic efficacy, even after controlling for place and each individual's age, gender, and education. Among participants in Indonesia, *P* is insignificant: simply attending meetings is strongly associated with individual participants telling interviews that they had more confidence in their civic efficacy after the last program meeting than before the first, regardless of place, the background of the person, or the intensity of their participation. Among participants in Tanzania, by contrast, the period of the program is not significantly associated with changes in perceived civic efficacy. Instead, changes in participants' confidence in their civic efficacy over the course of the program vary with how intensively participants engaged with the program. Those who participated intensively were significantly more likely to tell interviews that they were more confident in their civic efficacy at the end of the program than before the first, while participants who participated less were significantly more likely to describe lower civic

[36] The index is calculated on a 1-10 scale, as follows: 0-1 point for proportion of the meeting attended; 0-5 points for number of times the participant spoke distinctly, relative to the most verbose participant; 0-2 points for reporting having done an activity, relative to the most active participant; 0-2 points for volunteering to do an activity or being assigned one, relative to the most active participant.

efficacy. Figure 2 visualizes the expected change in the likelihood of a participant self-reporting each level on the 1-4 civic efficacy scale depending on the intensity of their participation. Those who were less engaged—those with a 3 or below on the 10 point *P* index—were more likely to have changed their response from completely to mostly able to improve their communities (4 to 3 on the 1-4 scale), while those who were more engaged were proportionately more likely to increase their response to completely able.

Figure 2 – Estimated Change in Civic Efficacy at Different Levels of Participation



The changes in Table 4 in participants' confidence in their civic efficacy over the course of the program must be interpreted cautiously. First, they are likely to partly reflect differences in individuals' interpretations of what the interviewer meant by being able to make improvements to their community. To allow for interpersonal comparisons on the same scale, interviewers also told participants a series of anchoring vignettes (Masset 2015; King et al. 2004) of three parents in a community like the participants' who are frustrated that their local teacher frequently misses class and often does not try hard to teach even when he does show up. One frustrated parent does nothing about the problem; a second tries to talk to the teacher about improving but the teacher will not listen and the parent thinks of going to village head for help but eventually gives up; a third enlists the support of the village head to talk to the teacher, who

subsequently makes noticeably more effort to teach well. Each respondent was then asked to rate the civic efficacy of each of the individuals in these stories on the same 1-4 scale. Their answers to these questions allow partial adjustments to make respondents' own self-ratings more comparable across individuals. Table 6 shows re-estimates of equation (1) modeled as a compound hierarchical ordinal probit. In Indonesia, the results are similar to the unadjusted model, but the association between increased civic efficacy and the period of the program is weaker. In Tanzania, the estimated change in civic efficacy is still 0 for those with a 3 on the 10-point *P* index; those who participated less described having lower civic efficacy at the end of the program and those who participated more described higher civic efficacy.

Table 6 – Civic efficacy: compound hierarchical ordinal probit estimates of equation (1)

	Both countries		*Indonesia*		*Tanzania*	
End of the program (*time*)	0.211	-0.397	0.544**	0.253	-0.666	-2.251***
	(0.175)	(0.330)	(0.252)	(0.464)	(0.504)	(0.858)
Participation Index (*P*)		-0.0326	0.0724	0.0175	0.448**	-0.350
		(0.0836)	(0.0656)	(0.106)	(0.207)	(0.277)
time x *P*		0.207**		0.0872		0.740**
		(0.0993)		(0.127)		(0.302)
Controls						
Age (thirds)	0.0573	0.0248	0.428*	0.397	-0.696	-0.438
	(0.159)	(0.163)	(0.252)	(0.262)	(0.458)	(0.348)
Education (quartiles)	0.279**	0.280**	0.247	0.247	1.385**	1.118**
	(0.135)	(0.138)	(0.157)	(0.160)	(0.707)	(0.538)
Gender (female)	-0.295	-0.247	-0.198	-0.163	-0.0491	-0.135
	(0.228)	(0.230)	(0.343)	(0.344)	(0.607)	(0.472)
Fixed effects for community	Y	Y	Y	Y	Y	Y

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Sample restricted to respondents in both meetings.
Both Countries: Respondents: 219. Communities: 31.
Indonesia: Respondents: 108. Communities: 15.
Tanzania. Respondents: 111. Communities: 16.
See Appendix C for estimates with alternative indices for participation.

Second, re-estimates of equation (1) substituting two alternative measures of participation—an index of participation derived from a factor analysis to represent the underlying intensity of participation, and the number of times each person spoke in meetings—are similar and in some models reflect relatively greater increases among those who were quieter in meetings. These estimates suggest that in Indonesia participation was associated with increasing confidence in civic efficacy, and offer some evidence that those who were most active were already relatively more convinced of their capacities so that the largest increases were concentrated among those who quieter and less active. However, when using vignettes to adjust for interpersonal differences in interpreting the interviewers' question, both *time* and the interaction with participation are insignificant in some models. (See Appendix C for re-estimates.)

Finally, even for those whose confidence in their civic efficacy did change over the program, these changes are only associated with the period of the program; they do not necessarily mean that participating in the program caused participants to be more confident in their civic efficacy. Changes in confidence may

have caused changes in participation in the program rather than the other way around: as some gained confidence, they participated more in the meetings or activities outside them, while others lost confidence and participated less. A third possibility is that these changes may also reflect factors other than those we controlled for that affected both participation and responses to interviewers' questions.[37]

Yet altogether, these interviews and the observations of participation in meeting discussions described above suggest from several perspectives that participants started the program with varying confidence in their civic efficacy but that most thought that they would be able to improve their community's maternal and newborn health care, and that a large majority remained confident after attempting the approach they had planned. Interviews and meeting observations also suggest that some participants seemed less than fully confident in their civic efficacy over the course of the program, and that in Tanzania they slightly outnumbered those who gained confidence. But observations of meeting discussions also suggest that in both countries, some participants who started skeptical had become optimistic by the end, consistent with them having discovered latent capacity they were skeptical of at the start. And interviews with individual participants suggest that particularly in Indonesia, many who described themselves before the program as less confident in their capacities to improve their communities had increased their perception by the end.

Later Reflections and Changes in Health Care

Focus groups two years after the program ended reflect whether and how participants remembered their efforts in hindsight and when the program was far enough in the past for participants to face fewer incentives toward socially desirable responses than they may have perceived before, during, or immediately after the program. These reflections suggest that even two years later, participants in almost all communities had memories of specific efforts they had taken to try to improve their care and of what had worked. They also suggest that the differing experiences between the two countries did not last once the facilitator was no longer holding meetings (Table 7).

[37] In particular, intensive ethnographic studies of how participants responded to and understood the program suggest that particularly at the beginning of the program, many participants were inclined to view their participation during the program as an investment that they hoped would lead to more resources (cite ethnography book).

Table 7 – Reflections on participation

Proportion of communities in which participants:	*Indonesia*	*Tanzania*
recalled		
specific activities that they had tried to improve their health care	97%	100%
at least one specific activity that they thought had been successful	93%	100%
at least one specific activity that they thought had been unsuccessful	39%	41%
at least one specific, tangible improvement as a result of their efforts*	41%	30%
thought that their activities overall had improved health care in their community	83%	95%
described specific activities to improve health care that they were still engaging in (individually or as a group)	48%	33%
said they were still meeting as a group	23%	26%
said their last meeting was		
within the last year	18%	26%
within the last six months	14%	25%
within the last two months	8%	14%
within the last month	4%	9%

Notes: * see Table 9.

In both, participants in a substantial minority of communities (39% in Indonesia and 41% in Tanzania) recalled lacking the resources, knowledge, or willingness to see their efforts through; remembered those with whom they tried to engage ignoring them, being unwilling to help; or otherwise described being unsuccessful in their efforts to fix problems with their health care. But in most, they also recalled at least one of the activities they had planned being successful. In a substantial proportion, they described activities they were still engaged in to try to improve their care, either individually or as a group. Reflecting on their efforts overall, participants in 83% of communities in Indonesia and 95% in Tanzania described at least some of their efforts as having improved their community's maternal and newborn health care (Table 7).

To what degree were these perceptions accurate? In Arkedis et al. (2019), we analyze interviews conducted also two years after the program ended with a recently pregnant women in each community about the quality and use of maternal and newborn health care in these communities and observations of their facilities, as well as in two hundred other randomly selected communities across the same four pronvices of Indonesia and Tanzania who were not offered the program.

On one hand, these interviews with recently pregnant women and observations of their facilities indicate that health care outcomes around pregnancy and birth had increased in communities who were offered the program. Compared to similar interviews and observations prior to the start of the program, they suggest that in the years since more pregnant women were receiving care and delivering at clinics and with skilled providers (Table 8). These measures also suggest a slight decline in the proportion of infants

who were underweight, although no change in the proportion of infants whose growth was stunted (less than 2 standard deviations below the median WHO Child Growth Standards) in Indonesia and a slight increase in stunting in Tanzania.

Table 8 - Change in Use of Health Care and Infant Health Outcomes

	Indonesia			*Tanzania*		
	2015	*2018*	*% change*	*2015*	*2018*	*% change*
Any antenatal care in first trimester	--	--	*	19%	23%	+4%
4 or more antenatal care visits	--	--	*	44%	53%	+9%
Birth with a skilled provider	78%	86%	+7%	56%	68%	+12%
Birth at a facility	55%	74%	+19%	56%	67%	+11%
Lack of stunting – Infants taller than 2 standard deviations below the median WHO Child Growth Standards	84%	84%	0%	74%	71%	-3%
Adequate weight – Infants weighing more than 2 standard deviations below the median WHO Child Growth Standards	84%	87%	+4%	91%	93%	+2%

Notes: Statistics calculated from household surveys of recently pregnant women in 200 communities in Indonesia and Tanzania who were offered the program * Question changed between baseline and endline surveys.

Yet these interviews with recently pregnant women and observations of their facilities, as well as focus groups with participants, also all suggest that their efforts were not on average sufficient to have added significantly to these increases in the use of health care or to these changes in health outcomes.

First, in a majority of communities in both countries, when participants were asked what they remembered of the effects of their efforts, including activities they had described as successful (Table 7), most of the improvements they recalled were general or vague, not specific, tangible results that they described as resulting specifically from them trying the approach they had planned in meetings. Table 9 lists both these general improvements as well as more specific, tangible changes that participants in these focus groups recalled.

Table 9 - Improvements participants recalled as a result of their efforts

Vague, general improvements	*Specific, tangible improvements*
Staff more accessible (excludes staff stays in facility/village)	An ambulance
Services now available 24 hours	A new oxygen tank
Better' equipment or drugs (unspecified)	A new building
General community awareness	New staff
Health worker attitude or performance	A new generator
Improved cost/affordability	New or improved road
Posyandu activity	New or improved rooms at the facility (such as delivery room, inpatient room)
Improved sanitation/clean water at facility (without specifics)	New information board or specific information that is now included on an existing board (cost, midwife phone number)
Claimed outcome: such as more women giving birth in the facility	No expired medicine at facility
Household toilets	A new toilet
Cooperation with facility	A service or arrangement for cleaning the facility
Collected data	Access to clean water
Vague improvement in "access"*	A new pharmacy
Vague renovations or improved building	A new or improved maternity waiting home
New outreach services	Electricity
	A new small birth clinic (*posyandu* in Indonesia)
	Distribution of birth preparedness stickers
	New beds
	A new registration counter
	New waste bins
	A new suggestion box
	Staff who now stay in or near the facility

Notes: All improvements that participants mentioned resulting from their efforts were first classified into the categories above, and then further distinguished by whether they described specific, concrete, tangible changes or more general or vague improvements. * "Vague" improvements in access include participants describing the road as "improved" or the ambulance as "available," rather than remembering that they got a *new* ambulance or road or specific improvement to their road.

Second, the approaches they tried to improve their care were not unusual. The program was designed to encourage participants to plan and try approaches that they thought would work in their contexts, and the kinds of efforts that they ended up trying were those that were common in the places they lived. When recently pregnant women were asked if they were aware of members of their community trying the specific kinds of activities in their communities that participants had described trying—such as whether members of their community had tried recently to improve their access to their local health facility, increase their ability to afford care, or improve the facility's infrastructure—few were significantly more common in the average community offered the program than in the average community who was not, and two-thirds of these were barely significant ($p < .1$).[38] In Tanzania, recently pregnant women were aware of

[38] In Indonesia, recently pregnant women were aware of significantly more attempts to create community savings groups, improve staff performance, and hygiene and cleaning campaigns (all $p < .1$). In Tanzania, recently pregnant women were significantly more aware of attempts to encourage them to visit the facility ($p < .05$), new complaint or suggestion boxes ($p < .05$), or attempts by community members to build or request new facilities ($p < .1$). Note that given the overall number of these kinds of approaches, one or more differences could be significant simply by chance.

significantly more of all of these kinds of activities in the communities who had been offered the program ($p < .01$), but the difference was small: 44 percent, 5.6 percentage points (0.1 standard deviations) more than in communities that were not offered the program. In Indonesia, there was no statistically significant difference in the number of these kinds of activities of which recently pregnant women were aware in communities who were and who were not offered the program (Arkedis et al. 2019).

Third, communities offered the program were also not unusual in seeing measurable improvements in their health care; the broader context was also of steady improvement. As noted in Section 3, this was a period during which improving health care was a high priorities for the national governments of both Indonesia and Tanzania, as for international donors and the broader international community. Both interviews with recently pregnant women and observations of their facilities indicate that in both countries, health outcomes in the average community offered the program were not significantly different than health outcomes in the average community who was not (Arkedis et al. 2019).

Altogether, these interviews and facility observations as well as focus groups with participants about the effects of their efforts all suggest that in the average community offered the program, the efforts of those who participated did not add significantly relative to existing efforts to improve community-level health outcomes.

One interpretation of the seeming disconnect between participants' perceptions of their efforts and comparisons of their health outcomes to communities where no program was offered is that among participants in the program who described themselves as having improved their health care, there was some degree of biased or exaggerated attribution: what participants described were not effects of their efforts.

Yet the interviews, focus groups, and observations all also support a second interpretation: that some of what participates tried led to changes to their care or access to it among their communities that—even if they either substituted for similar efforts that were happening already,[39] or had effects on measured community-level health outcomes that were insufficient on average to add consistently to existing efforts—were nonetheless noticeable and memorable to them and their community.

First, focus groups with participants two years after the program suggest that in roughly a third of communities—41 in Indonesia and 30 in Tanzania—when participants were asked why they thought their efforts had improved health care in their communities, what they described was not vague but rather specific tangible changes: new or improved buildings, staff, and ambulances; information boards and suggestion boxes; waste bins, new toilets, and clean water supplies; staff who lived at or near the facility; and many more (see Table 9). Nor were they the only members of their communities to notice these changes: in the 65 communities in which interviewers spoke with those who attended meetings and with others in the community, they heard in these interviews about similar changes in 26. In 17, similar changes

[39] This substitution effect is a common, longstanding conclusion about the effects of foreign aid as well; many studies have concluded that it is "fungible," substituting for existing efforts and thereby freeing up resources for other priorities (see, e.g., Mosley, Hudson, and Horrel 1987; Boone 1996; Kosack 2003).

were also observed in facilities or described in interviews with recently pregnant women two years after the program.[40]

Second, as noted (Table 7), these focus groups also suggest that participants in a substantial number of communities seemed to perceive their efforts as successful enough to keep them engaged long after the program ended. In nearly half of communities in Indonesia and 33% in Tanzania, participants described specific efforts they were still engaged in to try to improve their health care. Some described what they were doing individually, but in about quarter of communities in both countries, participants also said that they were still meeting as a group. In 14% in Indonesia and 25% in Tanzania they said they had met within the previous six months, and in 8% of communities in Indonesia and 14% in Tanzania participants said their last meeting was within the last two months.

Finally, in these focus groups, participants were also asked about whether they had experienced any personal benefits or costs from participating. In almost all communities, participants described their participation as personally beneficial (Table 10), and many of their reasons had to do with improvements they had made to their community's health care. Many described pride in new ambulances, nurses, new or improved facilities; in members of their community being treated better at the facility; in helping those who had been reluctant to get services before to go to the facility for care; in men accompanying wives more often to get care; or simply in having worked together to try to improve their communities.[41] Many also described gaining knowledge or experience speaking publicly about community issues.[42] Like their reflections on the results of their efforts, their reflections on their participation were not universally positive: some also described participating as a waste of time; others said that they should have been paid for their efforts; others described being treated suspiciously by traditional birth attendants or by neighbors who thought that they were being paid. But in 65% of communities in Indonesia and 74% of communities in Tanzania, not a single participant in the focus group described any cost or negative consequence from participating. In 94% of communities in Indonesia and all communities in Tanzania, more than three quarters of participants in these focus groups described their participation as personally beneficial.

[40] In forthcoming work we explore in more depth the experiences of participants who recalled their efforts as unusually successful, achieving several of the kind of tangible improvements in Table 8, in an effort to explain what made their experiences different.

[41] The motivation to help their community was also the most common reason that participants gave for why they had participated when interviewers asked them immediately after the program ended. Participants mentioned an altruistic motivation in 40 of the 41 communities in Indonesia and 18 of the 24 communities in Tanzania in which interviewers asked participants about their experiences immediately after the program ended.

[42] The ethnographic studies of participants experiences also support this interpretation (see xxx)

Table 10 - Reflections on Benefits and Costs of Participating

Proportion of communities in which participants:	*Indonesia*	*Tanzania*
said they had benefited personally from participating		
almost everyone (>75%)	94%	100%
anyone	99%	100%
said they had experienced costs or negative consequences		
almost everyone (>75%)	15%	19%
anyone	35%	26%
said that they were glad that they participated	97%	100%

Altogether these reflections two years after a facilitator had invited them to meetings, combined with observations of several of their meetings and interviews with them before, during, and after the program, offer additional evidence that for most who participated throughout the program the experience sustained or improved their perceptions of their civic efficacy. Their paths were far from uniform: in particular, participants in Tanzania, who began almost uniformly optimistic, were slightly more likely than not to have been disappointed in their progress over the course of the program itself. But in the years after facilitators were no longer holding meetings, participants in Tanzania had become almost as likely as participants in Indonesia to recall specific, tangible changes from their efforts; although fewer were continuing their efforts, slightly more than in Indonesia were doing so as a group; and participants in Tanzania had become just as overwhelmingly likely to describe their efforts overall as having improved their health care and as personally beneficial. Interviews with recently pregnant women in these and other communities where no program was offered, as well as observations of their health facilities, also indicate that participants' efforts were neither unusual nor sufficient on average to have added significantly to the broader improvements in community-level health outcomes in Tanzania and Indonesia over the time that the facilitator held meetings and in the years afterward. Thus some of participants' perceptions of efficacy may stem from attribution biases. Yet interviews, observations, and focus groups all also offer evidence suggesting that in a substantial proportion of communities, participants' efforts were strongly associated with at least some changes to health care that, whether they were causally or coincidentally associated with participants' efforts, were meaningful enough that 1) participants were able to describe in detail specific changes they had led to that were supported both by interviews with others in their communities and observations of their facilities, both during and years after the program, and that 2) participants were willing to continue their efforts years after the facilitator had left.

7. Conclusion

In 2015 and 2016, several thousand people in 200 randomly selected Indonesian and Tanzanian communities were invited to participate in a community-led scorecard program in which a facilitator offered information about their maternal and newborn health care and encouraged them to try to improve

that care. In this paper we draw from observations of several of these meetings in 81 of these communities, observations of individuals' participation in these meetings in 35, interviews in 65 with those who participated before and after the program as well as others in their community, and focus groups two years later in all 200, to ask how those who participated experienced this program and perceived their efficacy at improving their care. Each offers a perspective on participants' experiences of the program that can compensate to some degree for the social desirability, observer, recall, and other biases in the others. Altogether, they suggest that for most who participated, their efforts sustained or improved their perceptions of civic efficacy.

In almost every community, those who came to the facilitator's meetings were willing to engage in sustained and largely self-directed discussions and deliberations about problems with their maternal and newborn health care, as well as in civic activities they planned to try to improve that care. Their experiences of these efforts differed, and left some initially disappointed, particularly in Tanzania. But far more seemed in meetings to think that they were improving their health care, significantly more described being more confident in their civic efficacy after the program ended than before it began, and in the years afterward participants in almost all described their participation as beneficial to their community's health care and to them personally. Two years afterward, participants in a third of communities could still recall specific, tangible changes their efforts had led to, and in nearly a quarter they continued their efforts years after the facilitator left.

These observations, interviews, and later reflections do not suggest that the scorecard program encouraged universally engaged civic participation in every place where it was offered. Nor do they suggest that the changes participants and others recalled were sufficient, on average, to significantly improve community-level health outcomes. Instead, evidence from interviews with recently pregnant women about the quality and use of their maternal and newborn health care services and observations of their facilities indicate that two years after the program, many people in other communities were also trying approaches similar to participants in the program, and that measured health care outcomes in communities who were offered the program were not significantly different on average from communities where facilitators had not held meetings (Arkedis et al. 2019). These findings echo theory and other experimental studies described in Sections 2 and 3 that suggests that participation that is encouraged but not otherwise enabled with resources is not sufficient. For the people who had been willing to engage substantially and sustainably throughout this program, one possible implication is that after having attempted to improve their public services or convincing others to try, they would have been discouraged or disempowered by the experience.

Yet interviews with participants, observations of their meetings, and their later reflections on the experience rarely suggest this implication. Some were clearly discouraged when improving care proved harder than expected. But most, particularly in hindsight, found their participation beneficial and effective for improving their community's care. Instead, these interviews, focus groups, and observations support two other interpretations of the disconnect between participants' perceptions of their efforts and

comparisons of health outcomes in their communities to communities where no program was offered. For some, the difference may be the result of biased attribution: participants thought or told interviewers that they were responsible for more than they were, exaggerating the efficacy of efforts that had little effect. Second, when participants' efforts did lead to identifiable changes in health care, many of these changes, although noticeable and memorable to them and their community, had relatively limited effects on average health care outcomes or substituted for others' efforts that would have had similar effects. These implications suggest caution in generalizing from the responses and reflections of participants in this program to contexts in which access to quality public services is stagnating or declining rather than broadly improving.

Yet at least in the context of steady improvement in health care, interviews, observations, and focus groups also all suggest from a range of complementary perspectives that civic participation need not be organic or externally enabled with resources for those involved to perceive it as helpful and even empowering. These reflections and observations suggest instead that even in diverse, randomly selected communities in two countries on two continents, some people were almost always willing to participate and capable of efforts that they thought had improved a valued public service and that sustained or improved their perceptions of their civic efficacy, even though they were encouraged by a facilitator who offered them only information and space to discuss and plan but otherwise no other resources or authority, other than what they already had as members of their communities and citizens of their countries. And in a substantial number of communities, they suggest that participants were willing to continue their efforts long after the facilitator was no longer holding meetings and that their efforts led to specific changes to their care or access to it that were noticeable and memorable to them. To that extent, their attempts to improve the responsiveness and effectiveness of their maternal and newborn health care in living up to its promise sustained or improved their perceptions of their civic efficacy.

References

3ie. 2018. "Does Community-Driven Development Build Social Cohesion or Infrastructure?" Governance Working paper brief. London: International Initiative for Impact Evaluation.

Agrawal, Arun. 2005. "Environmentality: Community, Intimate Government, and the Making of Environmental Subjects in Kumaon, India." *Current Anthropology* 46 (2).

Ananthpur, Kripa, Kabir Malik, and Vijayendra Rao. 2014. *The Anatomy of Failure: An Ethnography of a Randomized Trial to Deepen Democracy in Rural India*. Policy Research Working Papers. The World Bank.

Andrews, Matt, Lant Pritchett, and Michael Woolcock. 2017. *Building State Capability: Evidence, Analysis, Action*. New York: Oxford University Press. https://bsc.cid.harvard.edu/building-state-capability-evidence-analysis-action.

Arkedis, Jean, Jessica Creighton, Akshay Dixit, Archon Fung, Stephen Kosack, Dan Levy, and Courtney Tolmie. 2019. "Can Transparency and Accountability Programs Improve Health? Experimental Evidence from Indonesia and Tanzania." Working Paper.

Baiocchi, Gianpaolo, Patrick Heller, and Marcelo K. Silva. 2011. *Bootstrapping Democracy: Transforming Local Governance and Civil Society in Brazil*. Palo Alto: Stanford University Press.

Bandura, Albert. 1977. “Self-Efficacy: Toward a Unifying Theory of Behavioral Change.” *Psychological Review* 84: 191–215.
———. 1982. “Self-Efficacy Mechanism in Human Agency.” *American Psychologist* 37 (2): 122–47.
Banerjee, Abhijit V, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. 2010. “Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India.” *American Economic Journal: Economic Policy* 2 (1): 1–30. https://doi.org/10.1257/pol.2.1.1.
Banerjee, Abhijit V, Rema Hanna, Jordan Kyle, Benjamin A. Olken, and Sudarno Sumarto. 2018. “Tangible Information and Citizen Empowerment: Identification Cards and Food Subsidy Programs in Indonesia.” *Journal of Political Economy* 126 (2): 451–91. https://doi.org/10.1086/696226.
Bauer, P.T. 1972. *Dissent on Development*. Cambridge: Harvard University Press.
Besley, Timothy, and Torsten Persson. 2010. “State Capacity, Conflict, and Development.” *Econometrica* 78 (1): 1–34. https://doi.org/10.3982/ECTA8073.
Björkman, M., and J. Svensson. 2009. “Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda.” *The Quarterly Journal of Economics* 124 (2): 735–69.
Boone, Peter. 1996. “Politics and the Effectiveness of Foreign Aid.” *European Economic Review* 40 (2): 289–329.
Cardoso, Fernando Henrique, and Enzo Faletto. 1979. *Dependency and Development in Latin America*. Berkeley: University of California Press. https://www.ucpress.edu/book/9780520035270/dependency-and-development-in-latin-america.
Dupuy, Kendra, James Ron, and Aseem Prakash. 2016. “Hands Off My Regime! Government Restrictions on Foreign Aid to Non-Governmental Organizations in Poor and Middle-Income Countries.” *World Development* 84: 299–311.
Easterly, William. 2006. *The White Man’s Burden: Why the West’s Efforts to Aid the Rest Have Done So Much Ill and So Little Good*. Penguin.
———. 2007. “Was Development Assistance a Mistake?” *American Economic Review* 97 (2): 328–32.
Farmer, Paul, Jim Yong Kim, Arthur Kleinman, and Matthew Basilico, eds. 2013. *Reimagining Global Health*. University of California Press. https://www.ucpress.edu/book/9780520271999/reimagining-global-health.
Ferguson, James. 1994. *The Anti-Politics Machine*. Minneapolis: University of Minnesota Press. https://www.upress.umn.edu/book-division/books/the-anti-politics-machine.
Fox, Jonathan. 2005. “Empowerment and Institutional Change: Mapping ‘Virtuous Circles’ of State-Society Interaction.” In *Power, Rights, and Poverty: Concepts and Connections*, edited by Ruth Alsop, 68–92. Washington, DC: World Bank.
———. 2007a. *Accountability Politics: Power and Voice in Rural Mexico*. New York: Oxford University Press.
———. 2007b. “The Uncertain Relationship between Transparency and Accountability.” *Development in Practice* 17 (4–5): 663–71.
———. 2015. “Social Accountability: What Does the Evidence Really Say?” *World Development* 72 (C): 346–361.
———. 2016. “Scaling Accountability Through Vertically Integrated Civil Society Policy Monitoring and Advocacy.” Working Paper. London: Making All Voices Count, Accountability Research Center.
Fung, Archon. 2006. *Empowered Participation*. Princeton: Princeton University Press. https://press.princeton.edu/titles/7762.html.
Fung, Archon, and Erik Olin Wright. 2003. *Deepening Democracy: Institutional Innovations in Empowered Participatory Governance*. Verso.
Gaventa, John. 2006. “Finding the Spaces for Change: A Power Analysis.” *IDS Bulletin* 37 (6): 23–33.
Grindle, Merilee Serrill. 1997. *Getting Good Government: Capacity Building in the Public Sectors of Developing Countries*. Cambridge: Harvard University Press. http://www.hup.harvard.edu/catalog.php?isbn=9780674354197.
———. 2007. *Going Local: Decentralization, Democratization, and the Promise of Good Governance*. Princeton, NJ: Princeton University Press.
Hahn, Rüdiger, Daniel Reimsbach, and Frank Schiemann. 2015. “Organizations, Climate Change, and Transparency: Reviewing the Literature on Carbon Disclosure.” *Organization & Environment* 28 (1): 80–102. https://doi.org/10.1177/1086026615575542.

Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Cambridge, MA: Harvard University Press.
Holland, Jeremy, and Florian Schatz. 2016. *Macro Evaluation of DFID's Policy Frame for Empowerment and Accountability*. London: ITAD. https://itad.com/wp-content/uploads/2017/06/EA-Macro-Evaluation-Technical-report-Dec16-FINAL.pdf.
Hsia, Renee Y., Naboth A. Mbembati, Sarah Macfarlane, and Margaret E. Kruk. 2012. "Access to Emergency and Surgical Care in Sub-Saharan Africa: The Infrastructure Gap." *Health Policy and Planning* 27 (3): 234–44. https://doi.org/10.1093/heapol/czr023.
Jenkins, Rob, and Anne Marie Goetz. 1999. "Accounts and Accountability: Theoretical Implications of the Right-to-Information Movement in India." *Third World Quarterly* 20 (3): 603–22.
Joshi, Anuradha. 2010. "Review of Impact and Effectiveness of Transparency and Accountability Initiatives: Annex 1 Service Delivery." Institute of Development Studies.
Joshi, Anuradha, and Peter P Houtzager. 2012. "Widgets or Watchdogs?" *Public Management Review* 14 (2): 145–62. https://doi.org/10.1080/14719037.2012.657837.
J-PAL. 2011. "Governance Review Paper: JPAL Governance Initiative." MIT.
Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2): 263–91. https://doi.org/10.2307/1914185.
Keck, M.E., and K. Sikkink. 1998. *Activists Beyond Borders: Advocacy Networks in International Politics*. Ithaca, NY: Cornell University Press.
Khagram, Sanjeev, Archon Fung, and Paolo De Renzio. 2013. *Open Budgets: The Political Economy of Transparency, Participation, and Accountability*. Brookings Institution Press.
King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, N.J.: Princeton University Press.
King, Gary, Christopher J.L. Murray, Joshua A. Salomon, and Ajay Tandon. 2004. "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research." *American Political Science Review* 98 (1): 191–207.
Kosack, Stephen. 2003. "Effective Aid: How Democracy Allows Development Aid to Improve the Quality of Life." *World Development* 31 (1): 1–22.
Kosack, Stephen, and Archon Fung. 2014. "Does Transparency Improve Governance?" *Annual Review of Political Science*.
Kruk, Margaret E, Anna D. Gage, Catherine Arsenault, and Keely Jordan. 2018. "High-Quality Health Systems in the Sustainable Development Goals Era: Time for a Revolution." *Lancet Global Health* 6 (11): PE1196-E1252.
Kruk, Margaret E, Hannah H Leslie, Stéphane Verguet, Godfrey M Mbaruku, Richard M K Adanu, and Ana Langer. 2016. "Quality of Basic Maternal Care Functions in Health Facilities of Five African Countries: An Analysis of National Health System Surveys." *The Lancet Global Health* 4 (11): e845–55. https://doi.org/10.1016/S2214-109X(16)30180-2.
Lieberman, Evan S. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99 (3): 435–52.
Lieberman, Evan S., Daniel N. Posner, and Lily L. Tsai. 2014. "Does Information Lead to More Active Citizenship? Evidence from an Education Intervention in Rural Kenya." *World Development* 60 (August): 69–83. https://doi.org/10.1016/j.worlddev.2014.03.014.
Lipsky, Michael. 1980. *Street Level Bureaucracy: Dilemmas of the Individual in Public Services*. Russell Sage.
Mansbridge, Jane. 1980. *Beyond Adversary Democracy*. Chicago: University Of Chicago Press. http://www.press.uchicago.edu/ucp/books/book/chicago/B/bo3618381.html.
Mansuri, Ghazala, and Vijayendra Rao. 2013. *Localizing Development: Does Participation Work?* Washington, DC: World Bank.
Masset, Edoardo. 2015. "Measuring Empowerment in Rural India Using Vignettes." *Journal of Development Effectiveness*, 1–11.
McGee, R., and J. Gaventa. 2011. "Shifting Power? Assessing the Impact of Transparency and Accountability Initiatives." *IDS Working Papers* 2011 (383): 1–39.
Mill, J.S. 1843. *A System of Logic*. London.
Ministry of Health and Social Welfare. 2008. "Human Resource for Health Strategic Plan 2008-2013."
Mosley, Paul, J. Hudson, and S. Horrel. 1987. "Aid, the Public Sector and the Market in Less Developed Countries." *Economic Journal* 97 (387): 616–41.

Moyo, Dambisa. 2009. *Dead Aid: Why Aid Is Not Working and How There Is a Better Way for Africa*. New York: Farrar, Straus, and Giroux.
OECD. 2008. *Paris Declaration on Aid Effectiveness (2005); Accra Agenda for Action (2008)*. Paris: OECD. http://www.oecd.org/dac/effectiveness/34428351.pdf.
Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115 (2): 200–249. https://doi.org/10.1086/517935.
Prebisch, Raúl. 1950. *The Economic Development of Latin America and Its Principal Problems*. New York: Economic Commission for Latin America, United Nations Department of Economic Affairs.
Pritchett, Lant. 2015. "Can Rich Countries Be Reliable Partners for National Development?" Essay. Center for Global Development. https://www.cgdev.org/sites/default/files/CGD-Essay-Pritchett-Rich-Countries-Reliable-Development-Partners.pdf.
Raffler, Pia, Daniel N. Posner, and Doug Parkerson. 2019. "The Weakness of Bottom-Up Accountability: Experimental Evidence from the Ugandan Health Sector." Working Paper.
Rose-Ackerman, Susan. 1999. *Corruption and Government: Causes, Consequences, and Reform*. New York: Cambridge University Press.
Sachs, Jeffrey. 2005. *The End of Poverty: Economic Possibilities for Our Time*. New York: Penguin.
Scott, James C. 1999. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven: Yale University Press.
Seawright, Jason. 2016. *Multi-Method Social Science: Combining Qualitative and Quantitative Tools*. Cambridge University Press.
Sen, Amartya. 1999. *Development as Freedom*. New York: Alfred A. Knopf.
Singer, Hans W. 1950. "The Distribution of Gains between Investing and Borrowing Countries." *American Economic Review* 40 (2): 473–85.
Stern, E., N. Stame, J. Mayne, K. Forss, R. Davies, and B. Befani. 2012. "Broadening the Range of Designs and Methods for Impact Evaluations." Working Paper 38, London: DFID.
Storeng, Katerini T., and Dominique P. Béhague. 2016. "'Lives in the Balance': The Politics of Integration in the Partnership for Maternal, Newborn and Child Health." *Health Policy and Planning* 31 (8): 992–1000.
Transparency for Development. 2016. "Baseline Report." Cambridge: Harvard Kennedy School. https://ash.harvard.edu/files/ash/files/baseline_report.pdf.
———. 2017. *Citizen Voices, Community Solutions: Designing Better Transparency and Accountability Approaches to Improve Health*. Ash Center Occasional Papers. Cambridge: Ash Center for Democratic Governance and Innovation.
Tversky, Amos, and Daniel Kahneman. 1991. "Loss Aversion in Riskelss Choice: A Reference-Dependent Model." *Quarterly Journal of Economics* 106 (4): 1039–61.
UN General Assembly. 2000. *United Nations Millennium Declaration, Resolution Adopted by the General Assembly*. Vol. A/RES/55/2. http://www.refworld.org/docid/3b00f4ea3.html.
———. 2015. *Transforming Our World : The 2030 Agenda for Sustainable Development*. Vol. A/RES/70/1. http://www.refworld.org/docid/57b6e3e44.html.
UNDP. 1990. *Human Development Report*. New York: Oxford University Press for the United Nations Development Program.
Vothknecht, and Sparrow. *PODES Infrastructure Census 2011: Report on Infrastructure Supply Readiness in Indonesia - Achievements and Remaining Gaps*. TNP2K - PSF.
WHO. 2017. *The Republic of Indonesia Health System Review*. 1st ed. Vol. 7. Health Systems in Transition. New Delhi: World Health Organization.
Woolcock, Michael. 2013. "Using Case Studies to Explore the External Validity of 'complex' Development Interventions." *Evaluation* 19 (3): 229–248.
World Bank. 2004. *World Development Report*. Washington, DC: World Bank.
———. 2018. "Mobilizing Communities for a Healthier Future: Impact Evaluation of Social Accountability Interventions in Uttar Pradesh, India." Washington, D.C.: World Bank.
World Economic Forum. 2015. "Why Health Clinics May Hinder Developing Countries' Efforts." World Economoic Forum. https://www.weforum.org/agenda/2017/06/why-health-clinics-may-hinder-developing-countries-efforts.
Yahya, Talhiya, and Mohamed Mohamed. 2018. "Raising a Mirror to Quality of Care in Tanzania: The Five-Star Assessment." *The Lancet: Global Health* 6 (11): PE1155-E1157.

Zacka, Bernardo. 2017. *When the State Meets the Street: Public Service and Moral Agency*. Cambridge: Harvard University Press.
Zulminarni, Nani, Valerie Miller, Alexa Bradley, Angela Bailey, and Jonathan Fox. 2018. "Movement-Building for Accountability: Learning from Indonesian Women's Organizing." Working Paper. Washington, DC: Accountability Research Center, Just Associates, and PEKKA.