World Development 112 (2018) 74–87

Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

Crowdsourcing accountability: ICT for service delivery

Guy Grossman [a], Melina R. Platas [b,*], Jonathan Rodden [c]

[a] University of Pennsylvania, 208 S. 37th Street (225 Stiteler Hall), Philadelphia, PA 19104-6215, United States
[b] New York University Abu Dhabi, Social Sciences Building, A5-145, P.O. Box 129188, Abu Dhabi, United Arab Emirates
[c] Stanford University, 616 Serra Street, Encina Hall West Room 100, Stanford, CA 94305-6044, United States

ARTICLE INFO

Keywords:
Africa
Uganda
Political communication
ICT
Government responsiveness
Public service delivery

ABSTRACT

We examine the effect on service delivery outcomes of a new information communication technology (ICT) platform that allows citizens to send free and anonymous messages to local government officials, thus reducing the cost and increasing the efficiency of communication about public services. In particular, we use a field experiment to assess the extent to which the introduction of this ICT platform improved monitoring by the district, effort by service providers, and inputs at service points in health, education and water in Arua District, Uganda. We find suggestive evidence of a short-term improvement in some education services, but these effects deteriorate by year two of the program, and we find little or no evidence of an effect on health and water services at any period. Despite relatively high levels of system uptake, enthusiasm of district officials, and anecdotal success stories, we find that relatively few messages from citizens provided specific, actionable information about service provision within the purview and resource constraints of district officials, and users were often discouraged by officials' responses. Our findings suggest that for crowd-sourced ICT programs to move from isolated success stories to long-term accountability enhancement, the quality and specific content of reports and responses provided by users and officials is centrally important.

1. Introduction

In many low-income countries, the long route of accountability—citizens holding service providers to account via their representatives in government—has been sidelined in favor of community monitoring programs (Andrabi, Das, & Khwaja, 2017; Banerjee, Cole, Duflo, & Linden, 2007; Bjorkman & Martina, 2009; Duflo, Hanna, & Rya, 2012). Circumventing governments that have been seen as difficult to reach, non-responsive, or ineffective, these programs have encouraged citizens to hold service providers directly to account. But they have yielded mixed results while also raising sustainability concerns (Banerjee, Banerji, Duflo, Glennerster, & Khemani, 2010). Concurrently, in the past decade, access to mobile-phones has increased exponentially (Fig. 1), prompting an interest in harnessing innovations in information communication technology (ICTs) to solve some of the most intractable development challenges (Peixoto & Sifry, 2017). In this paper, we return to a modified version of the older "long route" model of accountability, armed with new technology that dramatically reduces the cost citizens incur in accessing government officials.

Specifically, we explore the effects of a new ICT platform (U-Bridge), on the delivery of public services in the health, education and water sectors. The platform was designed to bridge the gap between citizens, who possess information on local service delivery problems, and subnational government officials, who have the authority to hold service providers to account and to allocate resources according to need. Implemented in Arua district in northwestern Uganda, U-Bridge allows citizens to report service delivery issues directly to district local government officials through a free and anonymous text-messaging system.

The U-Bridge program is based on the idea that participatory grassroots programs need not necessarily be designed to circumvent government actors. Instead, grassroots efforts can be harnessed to better inform government officials of service delivery deficiencies. This division of labor between those who are most informed (community members) and those who are best positioned to act upon the information (public officials), lies at the heart of U-Bridge and other recent "crowdsourcing" initiatives (Meier & Munro, 2010).

We assume that public officials have the means to improve at least some areas of service provision, but are constrained in part by a dearth of targeted information on local service delivery

* Corresponding author.
E-mail addresses: ggross@sas.upenn.edu (G. Grossman), mplatas@nyu.edu (M.R. Platas), jrodden@stanford.edu (J. Rodden).

https://doi.org/10.1016/j.worlddev.2018.07.001
0305-750X/

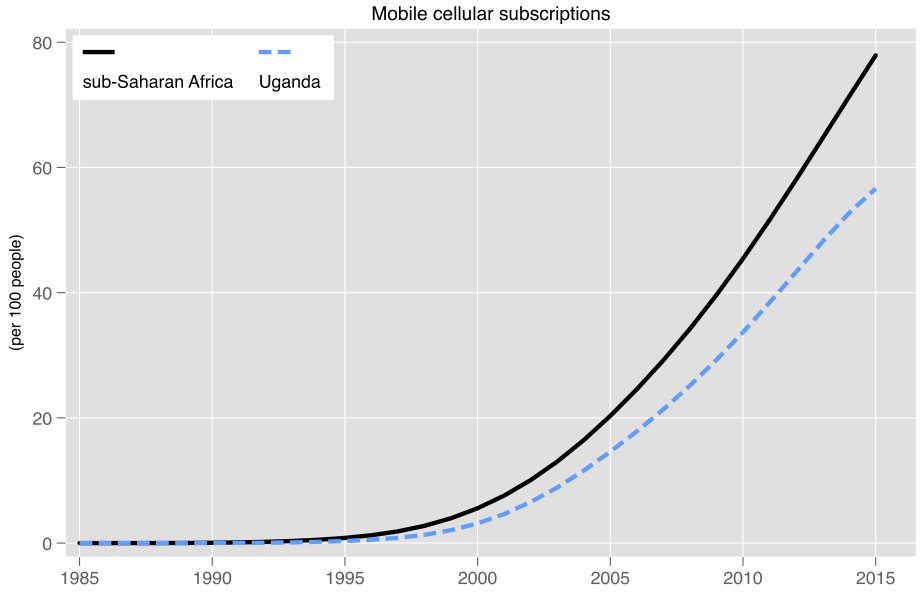

Fig. 1. Source: The World Bank; International Telecommunication Union, World Telecommunication/ICT Development Report and database.

problems. We further assume that citizens possess such information but are constrained in their ability to share it with government officials due to high monetary and social costs. To the extent that citizens believe that local government officials will be responsive, we expect citizens to use the ICT platform to make requests, complain, and raise concerns (Grossman, Michelitch, & Santamaria, 2017; Sjoberg, Mellon, & Peixoto, 2017). Knowing that local public officials have more and better information can increase citizens' expectations from government, which in turn, may incentivize public officials to exert greater effort (Gottlieb, 2016). Moreover, to the extent that service providers internalize the possibility of top-down sanctioning, relatively high usage of the platform may improve the efforts of service providers independently of public officials' actions. The overall effect should be an improvement in local governments' monitoring capacity and input targeting, as well as service providers' efforts.

To test the above theoretical expectations, we collaborated with GAPP, a non-government organization that has been working in Arua to improve the capacity of the district local government. GAPP developed the U-Bridge platform and implemented it in the district local government, training the district's political and technocratic arms on utilizing the system's functions. While all district residents could potentially contact Arua local district government via U-Bridge, only villages randomly selected by the research team were encouraged to use the ICT platform during the study period. Specifically, we constructed clusters of villages around Arua's 48 mid-level government health centers, and randomly assigned half of those clusters to the treatment group and half to the control. Each cluster consists of about 4–5 villages that are served by the same public health center and by at least one local public primary school. Villagers in treatment areas were informed of the new political communication channel through community meetings and an extensive door-to-door registration exercise that yielded more than 3000 registered (potential) users.

We collected outcome data on the performance of schools and health centers culled from Arua's line ministries (e.g., district health office) as well as from unannounced audits organized independently by the research team, the latter both to verify administrative data but also to collect data on measures not available in administrative data, such as absenteeism. We also collected administrative data on village-level requests for water parts, and the equipment received by villages to build and fix water service points. In addition, we conducted an endline survey in sixteen treatment villages to assess knowledge about, use of, and satisfaction with the U-Bridge service. For health and education, we examine three core outcomes: *monitoring* (e.g., inspections and calls to the facilities from district officials), *effort* (e.g., rates of staff absenteeism), and *input outcomes* (e.g., supplies of essential medicines, books and desks). We further collected data on health services utilization and on school test scores. We created indices combining these data sources for each type of outcome in each sector, by period (baseline, midline and endline).

Between August 2014 and November 2015, more than 11,000 messages were sent via U-Bridge to government officials in Arua district, though we estimate that only a quarter of these were directly relevant to service delivery. We find evidence of a positive treatment effect on education services (between one-fifth and one-quarter of a standard deviation in terms of the control group outcome distribution), but only in the first year. We also find suggestive but inconclusive evidence that U-Bridge may have had a positive effect on water parts and services over the study's duration. However, we do not find that U-Bridge had a discernible effect on education services beyond the first year, nor on health services at any point in time.

In the final part of this article, we discuss some of the reasons that may account for our mixed findings regarding the effectiveness of U-Bridge on service delivery outcomes, as well as the reasons why the modest effects of the program on education services were not sustained into the second year. Since there was relatively high uptake of U-Bridge compared to similar ICT platforms for political communication, at least in the first year, we can rule out the possibility that there was simply no demand for the service by citizens. Instead, we find that uptake of U-Bridge was highly uneven, concentrated in a handful of treatment villages, and that even in high-uptake villages, a plurality of citizens were not satisfied with the response of Arua district officials, which may have led to the drop in usage we observe overtime. Low satisfaction may have resulted in part from a mismatch in expectations about what constituted a useful and appropriate response to problems reported. For example, many responses by district officials referred users to lower-level local government or local institutions rather than directly following up on the problem reported. We also find that a considerable number of messages did not contain sufficient information that could be acted upon by local bureaucrats, or

concerned issues that could not plausibly have been addressed in a short time frame.

This paper heeds the call of Thompson (2008) and Heeks (2010), who point out the need for a new generation of empirical research on the extent to which, and conditions under which, ICT can contribute to positive development outcomes. One group of studies demonstrates that ICT interventions can lead to better information about, access to, and participation in markets (Muto & Yamano, 2009; Ogutu, Okello, & Oteino, 2014), as well as improved use of innovations in agricultural technology (Kiiza & Pederson, 2012; Larochelle, Alwang, Travis, Barrera, & Andrade, 2017). Our paper builds on a second group of studies that focus on whether ICT can improve governance and service provision (Bhatti, Kusek, & Verheijen, 2014). For instance, Kanyam, Kostandini, and Ferreira (2017) present evidence suggesting that mobile phone proliferation has led to a reduction in corruption, and Fu and Akter (2016) provide evidence that a mobile phone program improved the provision of public agricultural services in India.

Our paper also contributes to a related literature on information and accountability. This body of work is mostly concerned with the dearth of information in the hands of citizens. Much of the existing literature assumes, often correctly, that voters lack information about the quality of service provision that would otherwise allow them to better hold front-line service providers and/or politicians to account (Dunning et al., 2018). While there are certainly features of service quality that are difficult for users to observe, there are others that are visible—and perhaps uniquely visible—to users. Even if citizens posses reliable targeted information, they may be reluctant to confront service providers due to collective action problems inherent in community monitoring. Instead, in our study, citizens themselves are the providers of information to those in a better position to act upon that information: public officials.

Our study hints at the promise of ICT programs that focus on crowdsourcing as a route to better service provision, but perhaps more importantly, it also draws attention to some specific implementation challenges that can be addressed in the future.

2. Theoretical framework: bringing the state back in – with technology

Notwithstanding the almost ubiquitous adoption of political liberalization and decentralization reforms, many low-income countries exhibit the disappointing persistence of abysmal public services. Teachers and nurses are often absent without authorization (Chaudhury, Hammer, Kremer, Muralidharan, & Rogers, 2006), and many of those who do show up to work are unmotivated and exert low effort levels (Kremer, Duflo, & Dupas, 2011). Patients regularly receive incorrect diagnoses and treatments (Amin, Das, & Goldstein, 2008), and schools and clinics are often short of vital inputs (Glewwe, Kremer, & Moulin, 2009). Weak service provider efforts may be due in part to poor remuneration and resources, but they are also partly the result of poorly functioning local governments that lack the capacity and/or the will to adequately monitor front-line service providers and hold them accountable.

Following the publication of the 2004 World Development Report, *Making Services Work for Poor People*, numerous programs have sought to improve front-line services by focusing on the "short route" of accountability – empowering citizens to hold service providers accountable directly rather than indirectly via public officials (Kosack & Fung, 2014). This has coincided with a growing perception that local government officials are hard to reach and may have interests that do not align with those of their constituents (Bardhan, 2002). Short-route accountability models are based on two key assumptions: first, that communities can identify service delivery problems; i.e., that they posses information on both the quality of existing services and the level of services to which they are entitled; second, that communities can coordinate on their expectations from providers and on their own monitoring, and can do so in a sustainable manner.

Studies of programs designed explicitly to achieve short-route accountability, however, find mixed results. Alongside several notable successes (for example, Bjorkman & Martina (2009) and Andrabi et al. (2017)), other studies find no effect of community monitoring initiatives on service delivery and associated outcomes (Banerjee et al., 2010). Indeed, there is growing evidence that direct community monitoring is subjected to non-trivial collective action problems (The World Bank, 2016). Specifically, there is growing recognition of the power and status asymmetry that make rural villagers reluctant to confront service providers, partly due to fear of retribution (Kaawa-Mafigiri & Walakira, 2017).

Our research project revisits an old idea—bringing the state back in—while patching up holes in the logic of long-route accountability models with a new technology that reduces the costs of political communication and that targets non-elected officials in addition to elected leaders. Long-route accountability crowd-sourcing models start with the idea that the direct principals to which service provider agents answer are local government officials, not citizens. If a health worker is routinely absent, community members have no means of firing, transferring, or otherwise penalizing the health worker. They can only apply informal social sanctions, to which the health worker may or may not respond.

However, contacting public officials to demand that they hold service providers accountable has traditionally been (a) expensive and thus infrequent, and (b) perceived as potentially risky for citizens due to fear of retribution, especially for those residing in small communities. The modal way of contacting public officials still entails traveling in-person to the district headquarters. Due to poor roads and dearth of personal and public transit options, transportation costs in sub-Saharan Africa are notoriously high—at least twice those of the typical Asian country (Kessides, 2005). The crowd-sourced model of accountability explicitly addresses these twin problems. ICT platforms can link citizens directly to local government officials and allow free, unmediated communication as frequently as citizens wish. If user anonymity can be secured, personal risk can be dramatically minimized or eliminated.

Reducing the monetary and social costs of political communication, does not, however, guarantee that citizens will share information that is difficult for the local government to collect (Grossman, Michelitch, & Santamaria, 2017). Citizens will only use ICT platforms for service-related communication to the extent that they believe that message recipients are both resourceful and responsive. Unlike similar ICT platforms linking citizens to elected politicians, the primary recipients of messages in our study were local bureaucrats, although in practice both elected and unelected district officials had access to tablets and incoming messages from constituents. This was important in the Ugandan context, since bureaucrats are directly responsible for implementing public services and managing resources, and are mandated to promote, hire, and sanction service providers. Meanwhile, local politicians are tasked with monitoring bureaucrats, though they frequently lack the knowledge and resources to do so effectively. Thus, the U-Bridge program served as a sort of "medium" route of accountability, including elected politicians but more directly targeting local bureaucrats.

The crowd-sourced model assumes that public officials are interested in improving service provision, but are constrained by a dearth of reliable information on targeted problems. This may be a reasonable assumption in meritocratic bureaucracies that cultivate an esprit de corps among civil servants. It may, however, be

overly optimistic in patronage-ridden systems where citizens have low expectations from public officials. We argue that crowd-sourced platforms might be positioned to increase public officials' motivation by creating common knowledge around the reporting of service delivery problems. Under the status quo, citizens know that distant public officials lack reliable information on specific problems, which reduces citizens' expectations that officials will address service delivery deficiencies. Low expectations from citizens further reduces the pressure on public officials to perform (Gottlieb, 2016). By allowing citizens to easily report problems, ICT crowd-sourcing platforms may also increase citizens' expectations, since they now know that public officials know about problems, and public officials know that citizens know that public officials posses better and more accurate information. Such common knowledge—if created—can serve to increase citizens expectation, and in turn, public officials' incentive to address problems.

Importantly, if citizens choose to use the platform to report service delivery problems, and if public officials follow up on citizens' reports, then service providers might increase their efforts pre-emptively to avoid sanctioning. In this case, the *possibility* of reporting 'bad behavior' (rather than actual reporting) may sustain relatively high level of service provision over time. We thus designed our study to allow comparing both short-term (year-1) and longer-term (year-2) effects.

We expect that there are a set of preconditions that must be met for a program like U-Bridge to affect service delivery outcomes. These preconditions include:

1. Underlying demand for a means of communicating with local government;
2. Potential users do not fear retribution for reporting;
3. Potential users believe the local government has the mandate, capability, and interest in solving service delivery problems;
4. Messages contain *actionable* information – that is, information that is (a) new to the local government, (b) detailed enough to allow for a response, and (c) about a problem the local government has the resources and capacity to address;
5. Local government has the capability and interest to solve problems reported by users.

The data we present below allows evaluating the extent to which these preconditions were met, and whether U-Bridge was able to positively affect service delivery outcomes.

2.1. Domains of service delivery and hypotheses

Building on the above discussion, we expect that an ICT platform facilitating communication between citizens and district officials would affect three domains of service delivery: (1) monitoring, (2) effort, and (3) inputs, described below.

Monitoring: The first-order change we anticipate is that local government officials increase their *monitoring* of facilities. The district headquarters and thus district officials are often located quite far from facilities (in our study, mean distance between Arua town and study villages is 25 km as the crow flies, and longer by road), and have limited time and resources to spend physically visiting facilities. Though district officials generally know that there are problems with service delivery, as mentioned above, they commonly lack specific and timely information on delivery deficiencies. Thus, we expect that if local government officials hear complaints about a particular facility through the ICT platform, they will be more likely to exert effort on monitoring and supervising that facility, for example by making calls to or by visiting the facility.

Effort: We also expect an increase in effort of frontline service providers. Here we focus on activities that are under the direct control of individuals employed in the facilities: for instance, making requests to the district for additional goods and staff and conducting outreach activities. We also consider direct measures of effort in schools and health clinics, including absenteeism among teachers and health care workers. Increased effort at the facility level might emerge either as a response to increased monitoring from district-level officials, the expectation thereof, or the anticipation of greater direct local involvement in the monitoring of service provision.

Inputs: We expect that the more local government officials receive feedback from citizens through the ICT platform, the more they will become sensitive to facilities' needs and citizens' priorities, which could lead to input allocation improvements. In health clinics, these include the number of staff and supplies like drugs; for schools, such inputs include the number of teachers and school supplies; for water, these inputs include parts and repair services.

In sum, we hypothesize that access to an ICT platform for political communication will:

H1 increase **monitoring** in health and education.
H2 increase **effort** in health and education.
H3 increase the availability of **inputs** in health, education, and water.

3. Research design

To test the above hypotheses, we employ a field experimental research design. Specifically, we test the effect of encouraging randomly selected villages in Arua district in northern Uganda to report service delivery problems via U-Bridge — an open-source software package that opens a new channel of communication from citizens to local government officials.[1] Citizens are able to contact district officials via U-Bridge by sending a text message to a short-code number at no cost. District officials, in both technical and political positions, are equipped with 3G tablets that enable them to access the messages anywhere, provided they have Internet access. Once logged into the platform's dashboard, government officials can read tagged text messages from Arua residents and reply to messages from within the system.[2] Importantly, incoming text messages from citizens are anonymized such that government officials can only use a unique case ID (and not a phone number) to track the status of complaints over time.

3.1. The U-Bridge program

Implementing a platform for receiving information and for responding to incoming messages in the line ministries is a district-wide activity. The U-Bridge program also included three additional activities that were only implemented in treatment villages:

1. *Community meetings*: GAPP organized both an inception meeting and periodic community dialogue meetings in treatment areas. The inception meeting was used to introduce the ICT platform, and to provide attendees with information, based on codified service delivery standards, on the level and quality of services to which citizens are entitled. Community dialogue meetings took place on a quarterly basis, allowing citizens to interact directly with district officials that have attended the

[1] U-Bridge was developed by UNICEF Uganda and RTI International. RTI is the prime contractor for the United States Agency for International Development (USAID)-funded program, Governance, Accountability, Participation, and Performance (GAPP), of which U-Bridge is a part.

[2] GAPP employed a full-time intern responsible for training public officials in Arua on how to log onto the platform.

meetings and providing an overview of actions that the district has taken around service delivery. Community dialogue meetings were also intended to help create common knowledge between district officials and villagers, the importance of which is detailed in the theory of change above. Finally these meetings shared feedback with facilities about the messages that users had sent about service delivery problems. The first round of meetings was held by GAPP in the last quarter of 2014 as part of the launch of the U-Bridge service. The last round of community dialogue meetings took place in early 2016.

2. *User registration*: GAPP created a database of potential users, which entailed registering a phone number and basic demographic information.[3] Registration took place at community meetings, but also through a door-to-door registration drive implemented by the research team. While registration is not necessary for sending or receiving messages via the U-Bridge platform, it allowed GAPP to contact registered users with reminders, or to share information about the services to which they were entitled as Ugandan citizens.
3. *Periodic polls*: U-Bridge registered members were invited to respond to short periodic polls administered by the research team. The polls were conducted on weekends via a robocall system operated by VotoMobile, a Ghanian-based social enterprise. At the height of the program, in mid-2015, there were 3062 registered and verified U-Bridge users, who consented to respond to such polls.[4]

The U-Bridge program was rolled out in August 2014, and several adjustments were made to the program along the way. Perhaps most importantly, a few months into the program, it became clear that in the absence of a fully dedicated community liaison officer, district officials were unable or had insufficient time to manage the organization of the platform. Specifically, the filtering of relevant messages to the appropriate district official was time consuming. Filtering, which reduces the ratio of noise to signal, included deciding which messages are relevant and thus "worth" being forwarded to district officials as well as which office is most appropriate to deal with the request. In December 2014 the implementing partner GAPP hired a full-time staff member to filter messages and help district officials follow up on user messages. Thus, the platform would not have functioned as well in the absence of a third-party facilitating district officials' responses.

3.2. Randomization to treatment

Our study focuses primarily on two salient public services—health and education—and as such, the research team randomized treatment around places where these public services are delivered. While there are hundreds of government primary schools in Arua, there are only 48 mid-level government health centers. Mid-level health centers include the second and third tiers (Health Center II and III) in a four tier system.[5] We therefore use health centers as the unit of randomization. Specifically, we constructed 48 village-clusters: group of nearby villages that are serviced by the same public health center. We create blocks defined by the level of the health center (II or III) and randomized within blocks with equal probabilities. Thus, half of Arua's 48 clusters were assigned to treatment and half to control. All villages within a cluster are assigned to the same treatment group.

In each cluster, about 5 villages were selected to be included in the study. These villages included (a) the village in which the health center was located (primary village), and (b) three or four additional nearby villages. The selection of additional villages was determined by the presence of a government primary school, and the location of the village relative to other clusters. Whenever possible, we selected villages that had a government primary school within their administrative boundaries, and/or that were located near the health center. We excluded villages that were located close to other clusters in order to limit spillovers. In total, there are 131 villages in 24 clusters across Arua that were assigned to treatment, and 112 villages to the control group. The mean number of villages per cluster is 5; the minimum number of villages per cluster is 2 and the maximum is 9.

A balance test at the cluster level, derived from the 2014 Ugandan census, reveals that the treatment and control clusters are balanced on a number of dimensions commonly shown to affect service delivery outcomes and collective action, including ethnolinguistic fractionalization (ELF), share employed in the formal sector, poverty, and distance to the local government headquarters (Table 1). The clusters are slightly unbalanced with respect to the share of literate residents, but the difference in means is substantively small and probably not meaningful.

Although we do not have a census measure of phone usage, about 60 percent of our endline survey respondents report that either themselves (31 percent) or someone in their household (29 percent) owns a mobile phone. Two-thirds of respondents provided a phone number at which they could be reached, which may have belonged to a friend or relative outside the household. Mobile penetration is likely low compared to less rural parts of the country, but the majority of adults had access to a phone. Still, the fact that more than a third of respondents in the 16 treatment villages included in the endline survey did not have a phone suggests some limits to the use of this technology in promoting accountability, as well as highlights the question of whose voice is heard by district officials.

4. Data and estimation strategy

We measure outcomes in service delivery in the three domains outlined above using two primary data sources: (1) unannounced audits of schools and clinics conducted by the research team and (2) administrative data culled from the district education office (DEO), the district health office (DHO), the district planner, and the district water office, compiled by research assistants.

Both of these data sources have strengths and weaknesses. On the one hand, a major benefit of audits is that they are conducted by a source independent of the government or facility, and are therefore less subject (though not immune) to manipulation. On the other hand, some measures collected during the audit capture data from a specific day, producing somewhat noisy measures. Administrative data, by contrast, may be subject to manipulation and misreporting, but is collected more regularly. Notwithstanding this limitation, there is no reason to believe that biases in reporting from the facility level to the district level are correlated with treatment. In the analysis, we combine data from both the audits and administrative data into indices organized by domain.

Measures of monitoring include visits and phone calls to facilities from district officials, as well as inspection reports. Measures of effort include absenteeism, staff meetings, outreach clinics, and teacher engagement. Inputs include drug availability in health clinics and student supplies in schools. Health utilization includes variables such as the number of children immunized, maternity

[3] The process entailed texting "Join" to 8500, after which the sender would receive a series of messages to collect basic demographic information.

[4] The majority of users (2591) have entered the program through a door-to-door registration campaign that took place between October and November 2014 in treatment villages. An additional 251 users registered following GAPP's community meetings. 220 users registered with U-Bridge independently.

[5] We excluded Health Center IV and hospitals as these have different mandates, focusing on more specialized care.

Table 1
Balance test (cluster).

	Control Mean	Treatment Mean	Difference of Means	p-value for Difference of Means
Adult population (>= 16)	1460.292	1602.917	−142.625	0.464
	(116.970)	(153.786)	(193.215)	
Mean age	20.575	21.059	−0.484	0.066
	(0.197)	(0.166)	(0.257)	
Share Lugbara tribe	0.895	0.971	−0.076	0.162
	(0.052)	(0.010)	(0.053)	
Ethnic polarization	0.501	0.514	−0.013	0.836
	(0.049)	(0.041)	(0.064)	
Ethnic fractionalization (ELF)	0.060	0.049	0.011	0.669
	(0.019)	(0.017)	(0.025)	
Religion fractionalization	0.289	0.282	0.007	0.860
	(0.031)	(0.025)	(0.039)	
Share literate	0.595	0.636	−0.041	0.098
	(0.016)	(0.018)	(0.024)	
Mean education (0–4 scale)	1.166	1.182	−0.016	0.744
	(0.025)	(0.042)	(0.049)	
Share with secondary education	0.235	0.241	−0.007	0.784
	(0.011)	(0.021)	(0.024)	
Share employed	0.845	0.855	−0.009	0.722
	(0.019)	(0.018)	(0.026)	
Share employed non-agri sectors	0.226	0.255	−0.028	0.518
	(0.021)	(0.038)	(0.043)	
Poverty Index	−0.086	−0.119	0.033	0.468
	(0.023)	(0.039)	(0.045)	
Distance to Arua (kms)	36.708	28.083	8.625	0.162
	(5.090)	(3.309)	(6.071)	
N	24	24	48	

admissions, and health center patient attendance. For the water sector, we collect data on two variables, parts and services, which includes water parts distributed and water services rendered (for example, repairing a borehole) to a given village, and village requests, which is the sum of all water-related requests a given village makes to local government. Details of the variables included in service delivery indices can be found in the SI, Section 1, including descriptive statistics for all outcomes measured as well as the source of the indicator.

4.1. Unannounced audits

Unannounced audits were conducted in treatment and control schools and clinics at three points in time: baseline (June and July 2014), before U-Bridge was implemented; midline (June 2015), a year after the program launched; and endline (March 2016), almost two years since inception and after all GAPP meetings and all poll messaging had concluded. Via the audits, we collected information on all three domains: monitoring (e..g, communication with district officials), effort (e.g., absenteeism), and inputs (e.g., stock of supplies). Audits took place in each of the 48 health centers, and in 90 primary schools: 46 in control and 44 in treatment clusters. We also conducted audits in 44 schools located close to the border with but outside Arua (i.e, in neighboring districts). A much larger treatment effect with respect to "quasi-control" schools (compared to treatment effect in Arua proper), would suggest spillovers in Arua. By contrast, a smaller treatment effect with respect to "quasi-control" schools would suggest a SUTVA violation—greater attention to schools in villages assigned to U-Bridge comes at the expense of less attention to control schools, and thus may not be present when the system is scaled up.

4.2. Administrative data

Arua local district government collects regular reports from public facilities. Enumerators from the research team compiled data collected by the district in order to supplement data collection via the unannounced audits. Specifically, we assembled data at the district line ministries that can help determine whether more resources (and attention) have been allocated toward service provision in treatment communities; i.e., that have access to U-Bridge. For example, the DHO received monthly reports from each facility in which the facility reports the number of patients treated and procedures undertaken. The research team extracted performance trends directly from these reports. Both the DHO and DEO keep records of inspectors' visits in school and clinics; these data allow testing whether greater administrative attention is being directed at service units in treatment areas. We also assemble data on stockouts and staffing information from the DHO, and staffing information from the DEO. These data thus provide a secondary source, in addition to the audits discussed above, of absenteeism rates and stockouts.

4.3. Analysis

The key outcome variables are organized by sector (health, education, water) and domain (monitoring, effort, inputs).[6] We combine outcomes from each sector-domain into summary indices using two methods. First, following Kling, Liebman, and Katz (2007) and Casey, Glennerster, and Miguel (2012), we estimate mean treatment effect, which entails (1) recoding outcome variables so that higher values always indicate "better" outcomes; (2) standardizing those variables to allow comparability of effect magnitudes; (3) imputing missing values at the treatment assignment group mean; and (4) compiling a summary index that gives equal weight to each outcome component. The second method follows Anderson (2008), who recommends constructing the summary index at stage (4) as a weighted mean of the standardized outcome component, where the weights—the inverse of the covariance matrix—are used to maximize the amount of information captured by the index.

[6] In education we also explore student performance and in health, facility utilization.

Table 2
Example messages.

Type	Message context
Actionable	In [XXX] health Centre, they have given us staff but they are not at the work place
	We have problem of water source in [XXX] village
	In [XXX] village, [XXX] parish, [XXX] subcounty there is no road, no teachers in schools, and no staff in the health centres please help us from these situations
	Too much consumption of alcohol by the head teacher and some teachers has spoilt [XXX] primary school please help us
	We in [XXX] Health Center are suffering from lack of nurses we lost one of our nurses so can they help us with some nurses we are just left with one nurse here
	In our village [XXX], we have nothing like borehole. women go & fetch water for 1 km away from the village. How can government help our village?
	There are no drugs in [XXX] HEALTH CENTER. In which ways can U-Report help us?
	[XXX] Health Centre charges women for delivery at Health Unit and demand cash for visits
Relevant (not actionable)	Help us to eradicate poverty, please!
	those who monitor Government programs, are those children important or not?
	What do you think about Education
	I want to discus about education
Irrelevant	HAPPY XMAS FOR YOU
	Its pleasure greeting you, thank you
	I have good papers and commitment,punctuality,trustworth
	Where is the game won?

We measure all outcomes in two periods, short-term and longer-term. Short-term outcomes are measured approximately one year into the program, while longer-term outcomes are measured after the program activities, including village meetings, have concluded, close to two years after the program was launched. Following Banerjee et al. (2007) we report results for each time period separately. We estimate two types of models. First, following Bruhn and McKenzie (2009), we fit the following OLS regression (base) model, which is a equivalent to analysis of covariance (ANCOVA):

$$Y_{ijt} = \alpha + \tau T + Y_{ij0} + \eta_{ij} + \epsilon_{ijt} \quad (1)$$

where Y_{ijt} is facility i in cluster j at time t (short or longer-term), modeled as a function of the treatment indicator T, and the value of the dependent variable at baseline ($t = 0$).

Second, we reshape the data into long format and estimate a multi-level model (ML). In addition to accounting for the baseline level of the outcome in question and facility type, the multi-level model allows accounting for unobserved facility random effects.[7] Both OLS and ML models include block fixed effects for facility type (η_{ij}). For health centers, as noted above, we blocked randomization by health center type and thus add block indicators in the health estimation. We include fixed effects for four school types: Government-aided Catholic, government-aided Protestant, government-aided Islamic, or government non-affiliated. Schools that are religiously affiliated are usually founded and supported by religious institutions but are managed and funded by the government.[8]

We report all outcomes with and without a set of covariates, cluster standard errors at the level of village clusters (the level of randomization), and compute one-sided p-values. We apply the same criteria across sectors for the selection of covariates: we adjust only for covariates that are unbalanced across treatment and control, and following (Lin, 2013), demean these covariates and interact them with the treatment indicator.

5. Uptake

Prior to reporting the study's core findings, we begin by reporting overall uptake, which is not an experimental quantity. Data culled from the U-Bridge system suggests high levels of adoption in comparison to similar ICT programs studied elsewhere (Belcher, Lopes, Sjoberg, & Mellon, 2016; Grossman, Humphreys, & Sacramone-Lutz, 2016). Between August 2014 and November 2015 over 10,000 messages were sent via the platform. However, the content of the messages was highly variable in quality, and in the majority of cases, the message did not report a specific problem that could have plausibly been addressed by the local government. The vast majority of irrelevant messages were screened by the implementing partner and were never seen by district officials.

We hand-coded the messages sent by users in this time period into three types: actionable, relevant, and irrelevant. Actionable messages were those we determined to contain information about a specific problem in a specific location that could plausibly have been addressed by the district without further information. Relevant messages were messages about service delivery, but included messages that were vague or included insufficient information for local governments to respond. Finally, irrelevant messages did not contain content related to service delivery. Table 2 provides a sample of messages and their coding. Fig. 2 provides information on monthly (left-panel) and cumulative (right-panel) number of relevant and actionable messages overtime. The figure shows many people were willing to use the ICT system, but their efficacy in reporting problems—sending an actionable message–was low. The monthly number of messages peaked after about six months (May, 2015), and then dropped, though still registering over 100 monthly messages even 15 months after the program's launch.

6. Results

Our main findings capturing both short and longer-term effect of U-Bridge encouragement in health, education and water, are reported in Figs. 3 and 4. We report four sets of results for each unweighted mean index: ANCOVA and multi-level (ML) models, each estimated with and without covariate adjustment (labeled CA and NC, respectively). For robustness, we also report results in which one-tailed p-values of the base ANCOVA models (Eq. (1)) are derived from non-parametric randomization inference (SI, Figs. 1–5). Randomization inference (RI) is especially important to report when the experiment's sample size is relatively small (Young, 2017), since unlike parametric models, RI p-values do not rely on satisfying asymptotic assumptions. In addition, we

[7] Random effects rely on the assumption that the error term is uncorrelated with the key input variable (treatment indicator): an assumption we can make in lieu of the random assignment to treatment.

[8] We exclude private schools from the analysis. The distribution of school type is not significantly different across treatment and control.

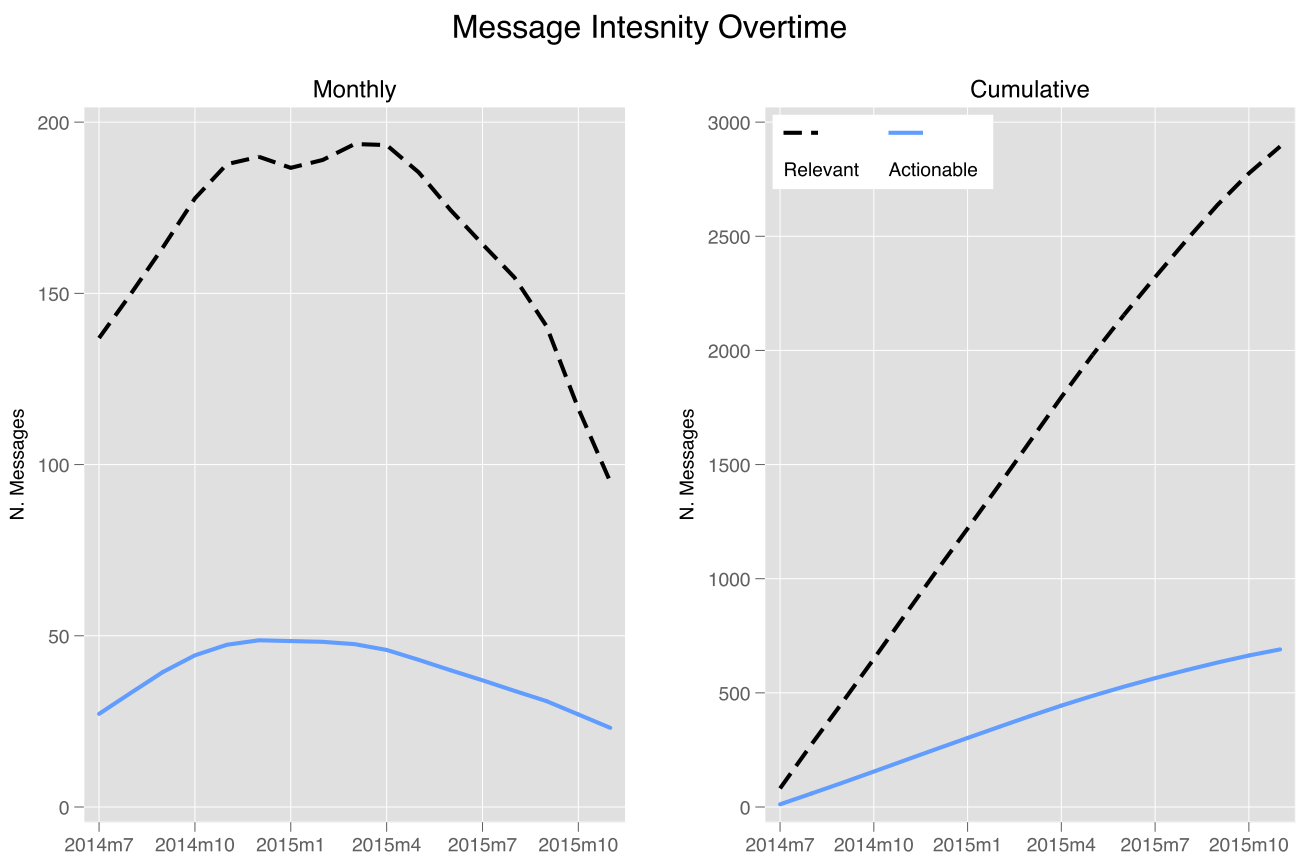

Fig. 2. The monthly (left-panel) and cumulative (right-panel) number of relevant and actionable messages over time. Lines are derived from locally weighted regression (lowess).

report findings for mean effect (weighted) indices, as per Anderson (2008), again for both ANCOVA and the multi-level model, with and without covariate adjustment. Main results and robustness checks in tabular forms, including results by all variables that comprise an index, are presented in the SI.

First, we find suggestive evidence of positive effects of the U-Bridge program for education outcomes, at least in the short-term, as shown in the top left panel of Fig. 3. Across specifications there is a consistently positive treatment effect on the monitoring and input indices, and, to a lesser extent, on the effort index. While the effect size is consistent across specifications (about one-fifth of a standard deviation), the treatment effect rarely reaches statistical significance at the conventional 0.05 level: one-tailed parametric p-values are 0.046, 0.133 and 0.070, and RI p-values are 0.067, 0.145 and 0.095, for education monitoring, effort and inputs, respectively in year 1.

The type of index used does not substantively affect findings, though the weakly significant results we report herein are even muted when employing the weighted indices (SI, Section 4.1). Even if they are genuine, the suggestive positive effects in the education sector, measured after the first year, were not sustained into the second year, shown in the top right panel of Fig. 3. While the coefficients on all three indices are positive, they are smaller than the coefficients in year one, and all are insignificant.

Second, we find little evidence that the U-Bridge program had an effect on any domain of health services in either the short or longer-term, shown in the bottom left (year 1) and right (year 2) panels of Fig. 3. The coefficients on the treatment indicator are negative across most health outcome indices. While we find a positive and significant effect on the health input index for the multi-level model, these results are sensitive to specification, and not observed for the ANCOVA model.

Turning to the water sector, where the post-treatment village-level data is aggregated across two years (2015–2016), we find positive coefficients for the treatment indicator on parts and services provided, but as shown in Fig. 4, these reach statistical significance only in the multi-level models. In order to gauge the substantive implications of the above coefficients, for the two outcome measures, SI, Fig. 13 provides information on the number villages by treatment group and period. Ultimately five more villages in treatment clusters had made requests for water parts and services in the post-treatment period (13 against 8 control villages), and five more villages received any water service in that period (9 treatment villages as compared to 4 control villages).

Together, we interpret these results as providing suggestive evidence of a positive treatment effect on educational outcomes in the short-term, particularly monitoring and inputs, but these effects appear to deteriorate over time. Our results for the water sector show a positive effect, but one that is never statistically significant. In health we find little evidence of any treatment effect. We briefly discuss several additional considerations in interpreting our results.

One concern is that our study was ultimately underpowered, and thus risks committing a *type-II error* – accepting the null hypothesis that there is no treatment effect when one exists in fact. In designing the study we conducted a power analysis to determine the number of clusters that would be required to detect a treatment effect. We anticipated that we would be able to detect a treatment effect of .25 standard deviations so long as the inter-cluster correlation within the 48 clusters is no larger than .085. In fact, the inter-cluster correlation was somewhat higher than we anticipated (around 0.18), which suggests that our minimum detectable effect size was around .35 standard deviations. This pro-

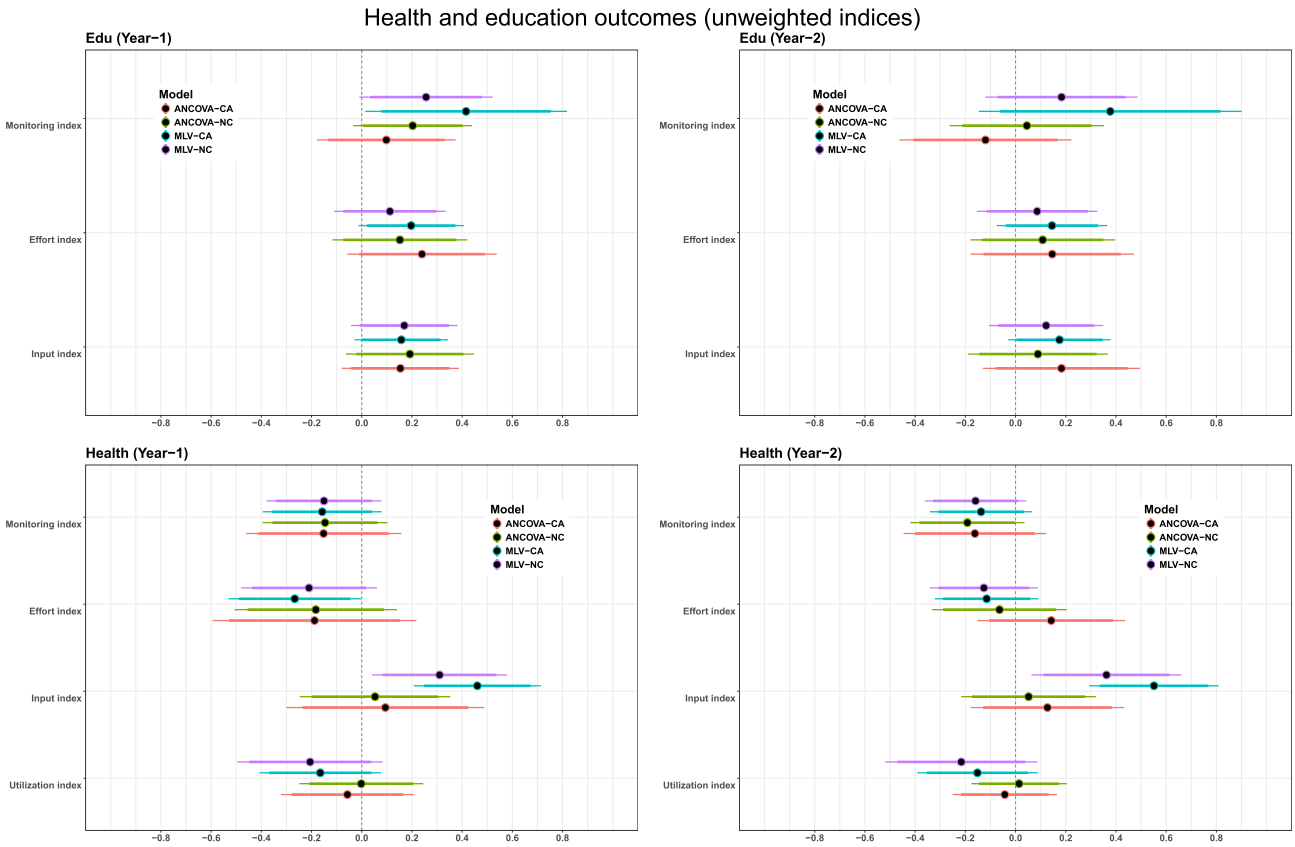

Fig. 3. The effect of U-Bridge on education and health services in years 1 and 2, ANCOVA and multi-level model with (CA) and without (NC) covariate adjustment.

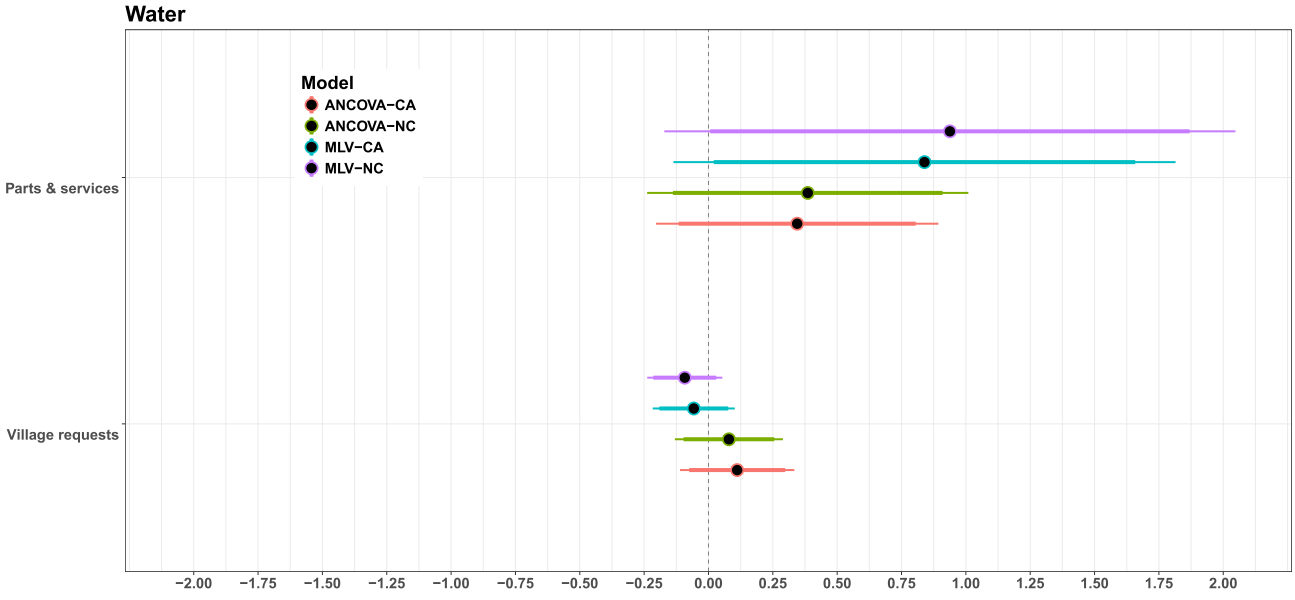

Fig. 4. The effect of U-Bridge on water: parts and services and village requests. Indices created using unweighted mean of standardized outcomes. Results presented with (CA) and without (NC) covariate adjustment. Models control for baseline level of the dependent variable and cluster standard errors at the cluster-level.

vides some evidence that we should be concerned about type II error in education. Thus rather than ruling out entirely the possibility of a treatment effect, we interpret the education and water results as suggestive. However, with the health sector, given the coefficients on the treatment effect are overwhelming negative, we are fairly confident there is indeed no treatment effect. We also note that, once we make adjustments for multiple hypothesis testing, there are no statistically significant effects for any index or sector (see the SI).

The final piece of evidence in support of a positive short-term treatment effect in education is an analysis comparing treatment facilities to facilities located in neighboring districts (quasi-control facilities). Recall that we collected data for some of our outcomes variables in a set of quasi-control facilities primarily to

assess the extent to which our treatment effects may be diluted by spillovers or enhanced by SUTVA violations. In the first scenario, district officials may responded to the program by increasing their efforts in all facilities in Arua – both treatment and control. In the second scenario, attention and resources toward treatment facilities are at the expense of control facilities. When we compare treatment facilities in Arua to quasi-control facilities in neighboring districts, we find positive and statistically significant effects for the monitoring and input indices in the education sector in year 1 at magnitude similar to the main treatment effects. These findings not only ameliorate concerns regarding both spillovers and SUTVA violations, but also provides us with further evidence that there was a genuine short-term treatment effect in education.

7. Discussion: exploring mixed results

In this section we explore our mixed findings. In particular, why is there suggestive evidence—albeit slightly below conventional levels of statistical significance—of positive effects of U-Bridge on education but not on health? Why do these effects, if in fact they exist, disappear by year two? In exploring these questions, we further evaluate the extent to which the preconditions for change, discussed above, were met.

Specifically, in addition to the audits and administrative data, we analyze the response messages sent by district officials, an endline survey of a subset of treatment villages, and qualitative interviews and focus group discussions with local government officials and U-Bridge users. The evidence at hand suggests that some but not all preconditions were met, and that any effects of U-Bridge on service outcomes are mostly due to piecemeal change. That is, it appears U-Bridge was able to resolve isolated incidents, but the evidence is not consistent with the creation of common knowledge that would allow for an equilibrium shift away from underperformance of service providers.

7.1. Why no effects in health?

Why did we find suggestive evidence of short-term effects in education and, to a lesser extent, in water, but none whatsoever in health? There are several possible explanations. First, it could be that there was less demand for improvements in health than in the other sectors, perhaps related to the fact that most citizens use schools and water points more frequently than health clinics. Relatedly, perhaps citizens are less able to identify problems in health services, as health requires a higher degree of comprehension or expertise (Peña-López et al., 2016). Second, local government may have greater control over the education and water sectors than health, and is thus better able to affect outcomes in the former. This could occur if, for example, health workers are more specialized and there is a lower supply than of teachers.

To address the first possibility, we examine whether there are differences in the number of messages sent by sector over time. Messaging intensity can serve as a proxy of citizens' demand for change. While many citizens interact with schools and water points on a nearly daily basis, they usually do not receive daily medical care. Our mixed findings may simply be a result of a difference in underlying demand for or the salience of health care. Fig. 5 suggests that this explanation is unlikely. Consistent with public opinion data on Ugandans' priorities over public services (Gottlieb, Grossman, & Robinson, 2016), the number of incoming messages (both relevant and actionable) exhibit similar patterns across public service type. There were not more messages about education and water than about health, nor are health-related messages less likely to be relevant or actionable.

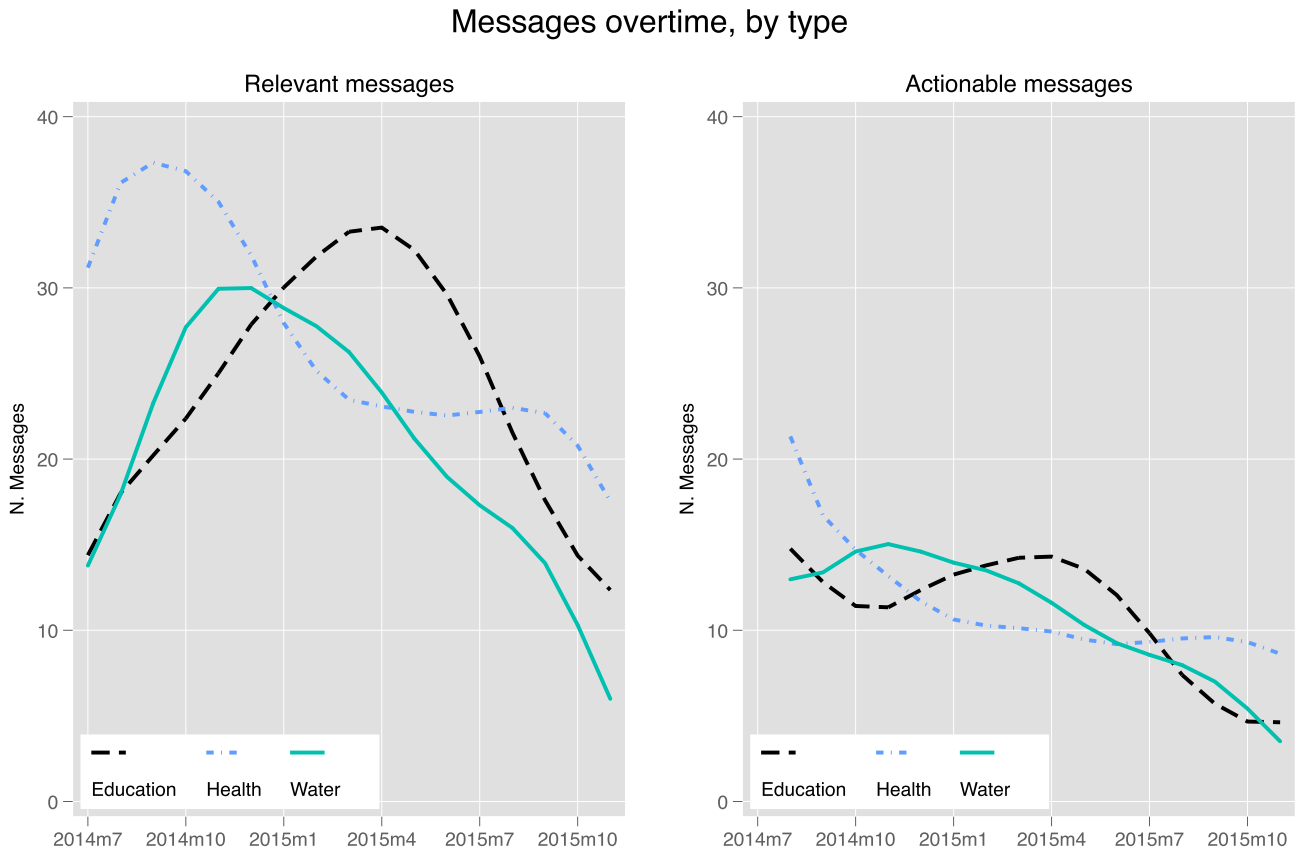

Fig. 5. The number of relevant (left-panel) and actionable (right) messages by service sector, overtime. Lines are derived from locally weighted regression (lowess).

Next we explore whether it is easier for the district to address complaints in the education sector than complaints in the health sector. This could occur if health inputs are relatively expensive, and because many aspects of health services, such as hiring doctors or stocking medicines, involve the central government, while education and water services are more decentralized. We thus plot the change in sectoral outcome indices between the baseline and the midline (year-1), against the number of relevant messages that we could attribute to specific village. The data at hand does not offer support to this explanation. At least in the case of Arua, the district education office is not more effective in translating messages into better outcomes, as compared to the district health office (SI, Figs. 14 and 15). More generally, we do not find evidence that positive effects in the first year for education were necessarily driven disproportionately by villages with a higher volume of messages (SI, Figs. 16 and 17).

7.2. Why don't the effects in education persist?

Perhaps the most exciting possibility of a program like U-Bridge is that it can generate common knowledge between service providers, citizens, and district officials about monitoring, thus generating incentives for better performance of teachers and headmasters in the long run. If common knowledge of this kind were being created in treatment clusters in Arua, however, we would expect that positive effects would persist or strengthen over time, even as the volume of messages decline. For example, the effect of U-Bridge could be sustained as teachers gained personal experience of being monitored, or perhaps heard that colleagues had been sanctioned for poor performance identified through the program, or that district education officials were visiting more often or asking harder questions.

Since the (suggestive) positive effects disappeared by the second year, it seems unlikely that this internalization took place. In our interviews with district officials, we learned of some examples in which information gleaned from messages was indeed used for disciplinary action, but this seems not to have been widespread enough to generate sustainable changes overtime. Moreover, in some cases, the disciplinary action taken was not very costly to service providers. For example, district officials cited the transfer of poor performing teachers from one school to another as evidence of the program's success. Such a transfer is not terribly costly for the teacher in question. Thus, it could be that the sanctions the district is able or willing to impose on poorly performing service providers simply are not costly enough to deter poor performance.

The results are consistent with a scenario in which initial excitement about and engagement with the program led to a temporary increase in monitoring activity from district officials, as well as short-term improvements in performance and physical inputs. However, it appears that this is difficult to sustain. As demonstrated above, the volume of messages declined substantially over time, and it is possible that service providers learned over time that the threat of vigorous monitoring and sanctioning was minimal. These possibilities are addressed in further detail below.

7.3. Why did the number of messages decline over time?

Although we do not find a significant positive relationship between the number of messages sent and a change in outcomes, it is still worth exploring the question of why the number of messages sent declined over time. First, it could be the case that after an initial registration drive and encouragement to use the system, which drove initial usage, users found the district government's responses unsatisfactory or ineffective in fixing the problems they had reported. Then, discouraged, they stopped using the platform.

A second possibility is that people forgot about U-Bridge. U-Bridge was nowhere near as interactive or entertaining as the most globally successful applications. Users would have to remember to use it, and remember the shortcode required to submit a message. A third possibility is that the initial messages, combined with the quarterly community meetings, satisfied users needs, such that there were fewer problems they wanted to report.

In order to learn about citizens' attitudes and knowledge of the U-Bridge system, we conducted a number of interviews and focus groups with users and government officials. In addition, we conducted an endline survey in spring 2016, interviewing all available adults (3192 respondents) in sixteen treatment villages; eight with high uptake of U-Bridge, and eight with low uptake. This exercise included a set of basic demographic questions, a set of questions that allow us to construct individuals' social network, and questions on knowledge and use of U-Bridge. This survey allows us to probe questions about variable uptake of U-Bridge—which we address in a companion paper (Ferrali, Grossman, Platas, & Rodden, 2018) —as well as to assess the extent to which assumptions required for U-Bridge to affect service delivery outcomes were met. Specifically, the endline survey sheds further light on the extent to which citizens expect that the district government has the capacity to affect service delivery, and their experiences in using U-Bridge.

The endline survey reveals that first, despite having had an intensive door-to-door registration campaign and a series of community meetings about the program, less than one third (31%) of respondents reported having heard of U-Bridge. Of those, 15 percent reported that they had sent at least one message. Thus, only a minority of potential users had in fact used the program, making it impossible to create the kind of knowledge that might sustain an initial response.

Further, 35 percent of users reported that they were somewhat or very unsatisfied with the district's response to their message (compared to 39% that were satisfied, with the remaining neither satisfied nor unsatisfied). Most dissatisfied users reported that they never heard back from district officials. Attitudes toward the program were not entirely negative, but neither do they indicate an overwhelming endorsement of the program.

On the other hand, 62 percent of users said they heard back from the district most or all of the time after sending a message, and the vast majority of respondents, 84 percent, stated that they believed the local government was "somewhat capable" or "very capable" of improving public service delivery. Thus, it is unlikely that users stopped sending messages because they came to believe district officials were unable to address service delivery problems. But perhaps despite hearing back from district officials, users did not see much change on the ground, and thus came to believe the system would not solve the problems they had hoped it would. The drop in messaging after six months, combined with low satisfaction rates at endline, is consistent with users becoming disillusioned over time.

7.4. How did officials respond to the program?

Finally, we turn to explore how government officials responded to the program. Did they find it helpful? Did it provide them with new information? Did they provide regular and informative responses to users? Did they use the information from messages they received to solve problems? To answer some of these questions, we conducted a focus group with district officials in late 2015 and conducted a second round of interviews in mid-2017. These discussions suggest that district officials were generally positive about the program, and pointed to specific instances in which they received new information that they were then able to act

upon. Some illustrative comments from these focus groups are in the Appendix.

Input provided from district officials shed light on the kinds of outcomes services such as U-Bridge could plausibly have affected and how officials perceived the program. First, district officials reported many requests that could not have been implemented in a short-time period, and others that were unrealistic given their budget implications. Second, a non-trivial number of individuals used U-Bridge to report problems not necessarily related to service delivery, such as corruption and crime. Third, district officials did report receiving at least some information they did not previously have, and that they subsequently acted upon. Thus, there is evidence U-Bridge did help officials to resolve at least some service delivery problems. Perhaps these were simply few and far between in comparison to the problems that the district could not address, or that were beyond the purview of this study.

An analysis of responses given by district officials from the start of the program through August 2015 suggests that officials often referred users to lower-level local government or other local institutions or bodies. For example, in response to concerns reported about the number of latrines, discipline of teachers, quality of teacher housing, and misuse of school funds, the district education office frequently referred users to consult the Parent Teacher Associations (PTAs), School Management Committees (SMCs), or head teachers. Indeed, these institutions may be the first point of call for parents, but it is possible that the reason users were reporting directly to the district was because efforts at lower levels to resolve problems had failed. It is possible the disappointment expressed by users was in part a result of a mismatch in their expectations about the type of response that would be provided by district officials and the response actually provided.

These insights yield potential lessons for future ICT programs like U-Bridge in settings with multiple overlapping layers of local government. Officials at the level of government receiving the messages might respond—perhaps truthfully—that the solution to the problem lies most clearly within the purview of another layer of government. But such responses are unlikely to solve the problem or satisfy the sender.

Together, we can rule out some possible explanations for a lack of lasting impact, while providing suggestive evidence of others. Many but not all of the preconditions for the efficacy of U-Bridge were met. First, we can rule out the possibility that there is no or low demand for a program like U-Bridge. Uptake was high in comparison to similar programs, and this was despite the fact that the program took place in a poor rural district with low mobile phone penetration. We expect uptake would be even higher in areas with higher mobile penetration especially if combined with a more extensive outreach campaign (using, for example, radio announcements). Second, we have evidence that many of the assumptions about the conditions under which citizens would use U-Bridge were met: people believed the district had capacity to respond, highlighted the importance of the anonymity of the program, and do not seem to have been free-riding, at least with respect to messages sent by friends and peers (Ferrali et al., 2018). Third, the evidence suggests that the content of messages was often insufficient to allow district officials to respond, and perhaps that even when messages were actionable, the district did not have the resources to respond to many of them, at least in the short-term. This precondition—that messages are actionable—was therefore only partially met, at best.

Self-reports from district officials as well as logs of outgoing messages from district officials suggest that the cases in which officials were directly involved in resolving service delivery problems were relatively few. More often, officials referred users to other institutions or individuals. We further find no sustained evidence of an increase in the monitoring activities by district officials that we measured, such as calls or visits to facilities, though it is conceivable that district officials employed forms of communication or monitoring that were not recorded, such as calling the subcounty local government.

We thus cannot rule out the possibility that district officials had insufficient incentives to act upon all of the information they received. The primary recipients of messages were bureaucrats holding unelected positions. We had positive and encouraging interactions with them in our numerous visits to the district headquarters, and as noted, they seemed interested and enthusiastic about the program. Nevertheless, we cannot say definitively whether a lack of incentives was consequential in this particular case. The program may also have been more effective if lower levels of local government, such as subcounties, were involved in receiving and responding to messages.

8. Conclusion

Circumventing government actors has become a popular strategy to improve service delivery in recent years, as researchers have tested whether the short route of accountability—empowering communities to directly monitor providers—can be effective in improving services, particularly in the context of rural and low-income settings. A key barrier in the long route of accountability has always been the cost of sharing information between service users and government officials responsible for overseeing providers of those services. We evaluated a program designed to overcome this barrier by providing a free and anonymous text messaging service through which citizens could quickly and cheaply share information about service delivery problems they witnessed on the ground.

We examined the extent to which the U-Bridge program, operating in Arua district in northern Uganda, could improve monitoring, effort, and inputs in health, education and water. Over the course of the period for which we have data, over 10,000 messages were sent, and district officials reacted positively to the program. While anecdotes of service improvements were plentiful, when we compared a variety of service outcomes in treatment villages—where the program was actively introduced and facilitated and where service requests were made—with a group of control villages, we do not find evidence of significant, sustained improvements in a variety of outcomes. We find suggestive evidence of short-term improvements in the education sector, but these were not sustained in the second year of the program. We also find only some suggestive evidence of greater activity in the water sector.

We evaluate a variety of potential explanations for our failure to identify a robust effect of the program on service delivery. We find that most of the preconditions for program success were met. We are able to rule out the possibility that there was no demand for such a program, that citizens did not believe the district would respond, or that potential users feared retribution.

Rather, we found that, first, despite relatively high levels of uptake compared with other settings, knowledge about the program was relatively low in treatment areas. Second, those that did send messages often sent messages that did not contain information the local government could act upon. Further, even when information was actionable, the district did not always have the resources or ability to respond, at least in the short-term, and this may have been frustrating for some users. Third, some messages that were actionable were not related to service delivery, such as crime, and thus their impact was not measured. Fourth, users who had sent actionable messages were frequently referred to lower-level local government or institutions, which they may or may not have contacted and which may or may not have been responsive. Indeed, it is entirely likely that the reason users sent

a message through U-Bridge was because local institutions had failed or users feared retribution from service providers. A related concern is that in a multi-layered decentralized system, a direct line of communication from citizens to one layer of government might simply allow officials at that layer to shift responsibility or blame for poor service provision to other layers.

Perhaps because so many messages did not result in problem-solving on the ground, user satisfaction with the program was relatively low, and likely declined over time. It is possible that as a result of the decline in messages or decline in the salience of the program in communities and government facilities, the already weak short-term treatment effects were not sustained over time.

A program like U-Bridge can provide information that is extremely useful in specific instances. For example, if a bridge is washed out in a storm, a roof is blown off a classroom, a teacher is abusive, or if there is a disease outbreak, word will probably eventually travel to the district headquarters. But a program like U-Bridge allows the information to travel to district officials much more quickly, potentially facilitating a quicker response. In this respect, such a platform can function much like a hotline. This kind of piecemeal impact, however, differs considerably from a situation in which the program ushers in a new equilibrium in monitoring and accountability, such that local service providers– when tempted to request extra fees from users or take the day off—are deterred because of an enhanced concern that monitoring will facilitate sanctioning. We are confident that U-Bridge achieved the first goal in some instances, but are unable to find evidence for the second.

Several policy implications emerge from this work. First, there appears to be significant demand for programs like U-Bridge, and thus it is worth investigating the ways in which such programs can be improved to better affect the intended outcomes. Second, such programs are unlikely to bring about a substantial change in the incentives of service providers, and hence discernible longer-term improvements in service provision, unless large numbers of people learn about the program and learn how to send actionable messages that focus on service provision outcomes that governments have the resources and ability to improve. Third, it is possible that programs like U-Bridge will be more effective if they target specific sectors or require users to select among a pre-established set of issue categories, rather than allowing users to send open-ended messages. Finally, even if citizens report actionable problems, local government officials must have the interest and mandate to resolve them, and coordination problems between layers of local government must be resolved.

Conflict of interest

None.

Acknowledgments

Data collection was conducted by Innovations for Poverty Action Uganda and Hatchile Consult, Ltd. The project was approved by the Uganda National Council for Science and Technology (#SS 3266), Office of the President in Uganda, Mildmay Uganda Research Ethics Committee (#REC REF 0204-2015) and Institutional Review Boards at Stanford University and the University of Pennsylvania. Funding for the project was provided by the United States Agency for International Development (USAID) and the Stanford Institute for Innovation in Developing Economies (SEED). A pre-analysis plan is registered at EGAP (ID 20160819AA). We would like to thank Jon Helfers, Maximillian Seunik, Zachary Tausanovich, Areum Han, Hardika Dayalani, Siyao Li, and Mitchell Goist for research assistance at various stages in the project, as well as Moses Banduga and Dickens Anguzu at GAPP. We also thank our research partners at Social Impact, and our implementing partners: GAPP, UNICEF Uganda, and the Uganda USAID mission. We also thank participants at workshops and presentations at the International Food Policy Research Institute (IFPRI), Carnegie Mellon University Africa, African Studies at New York University Abu Dhabi, the Center for Experimental Social Sciences at Nuffield College, TICTeC (Lisbon), Blavatnik School of Government at Oxford, and the World Bank's Governance Global Practice Community of Practice Citizen Engagement. Most importantly, we thank the Arua district local government for their interest and collaboration in implementing the project, and the thousands of residents of Arua district who participated in the U-Bridge program and our research activities.

Appendix

Sample of comments from district official focus groups

- *"U-Bridge allows for responses in near real-time. Example: [date] Community Dialogue Meeting showed that the In-Charge of [XXX] Health Centre had been absent for several months. Following the meeting, the DS reported to both the CAO [Chief Administrative Officer] and DHO [District Health Officer]. This resulted in district action and the In-Charge returned to his post, apologized and explained his absence to constituents."*
- *"While we remain constrained by resources, the platform improves the ability of councilors to reach out to constituents."*
- *"The platform has catalyzed how quickly the district responds to issues. For instance, community members reported that the Youth Chairperson was embezzling millions of shillings. They communicated to the district and police action was taken immediately."*
- *"Some messages require drastic changes in budgetary allocations. These are the hardest to respond to (e.g. "there are not enough health workers") or, those messages that require action at the central level."*
- *"U-Bridge also allows the District to broaden our mandate more fully. For instance, issues of domestic violence, child abuse, family matters, etc. can also be addressed. The U-Bridge allows district officials to refer issues more effectively (e.g. to the police)."*

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.worlddev.2018.07.001.

References

Amin, S., Das, J., & Goldstein, M. P. (2008). *Are you being served?: New tools for measuring service delivery*. Washington D.C.: World Bank Publications.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association, 103*(484), 1481–1495.

Andrabi, T., Das, J., & Khwaja, A. I. (2017). Report cards: The impact of providing school and child test-scores on educational markets. *American Economic Review, 107*(6), 1535–1563.

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

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics, 122*(3), 1235–1264.

Bardhan, P. (2002). Decentralization of governance and development. *The Journal of Economic Perspectives, 16*(4), 185–205.

Belcher, M., Lopes, C. A., Sjoberg, F. M., & Mellon, J. (2016). MajiVoice Kenya–better complaint management at public utilities. In *Civic Tech in the Global South* (pp. 179). The World Bank.

Bhatti, Z. K., Kusek, J. Z., & Verheijen, T. (2014). *Logged on: Smart government solutions from South Asia*. World Bank Publications.

Bjorkman, M., & Svensson, J. (2009). Power to the people: Evidence from a randomized field experiment on community-based monitoring in Uganda. *Quarterly Journal of Economics, 124*(2), 735–769.
Bruhn, M., & McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics, 1*(4), 200–232.
Casey, K., Glennerster, R., & Miguel, E. (2012). Reshaping institutions: Evidence on external aid and local collective action. *Quarterly Journal of Economics, 127*(4), 1755–1812.
Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., & Rogers, F. H. (2006). Missing in action: Teacher and health worker absence in developing countries. *The Journal of Economic Perspectives, 20*(1), 91–116.
Duflo, E., Hanna, R., & Rya, S. P. (2012). Incentives work: Getting teachers to come to school. *The American Economic Review, 102*(4), 1241–1278.
Dunning, T., Grossman, G., Humphreys, M., Hyde, S. D., McIntosh, C., & Nellis, G. (2018). *Metaketa I: Information, accountability, and cumulative learning*. Cambridge University Press.
Ferrali, R., Grossman, G., Platas, M., & Rodden, J. (2018). *Peer effects and externalities in technology adoption: Evidence from community reporting in Uganda*. Working Paper.
Fu, X., & Akter, S. (2016). The impct of mobile phone technology on agricultural extension services delivery: Evidence from India. *Journal of Development Studies, 52*(11), 1561–1576.
Glewwe, P., Kremer, M., & Moulin, S. (2009). Many children left behind? Textbooks and test scores in Kenya. *American Economic Journal: Applied Economics, 1*(1), 112–135.
Gottlieb, J. (2016). Greater expectations: A field experiment to improve accountability in Mali. *American Journal of Political Science, 60*(1), 143–157.
Gottlieb, J., Grossman, G., & Robinson, A. L. (2016). Do men and women have different policy preferences in Africa? Determinants and implications of gender gaps in policy prioritization. *British Journal of Political Science*, 1–26.
Grossman, G., Michelitch, K., & Santamaria, M. (2017). Texting Complaints to Politicians: Name Personalization and Politicians' Encouragement in Citizen Mobilization. *Comparative Political Studies, 50*(10), 1325–1357.
Grossman, G., Humphreys, M., & Sacramone-Lutz, G. (2016). Information technology and political engagement: Mixed evidence from Uganda. Working paper.
Heeks, Richard (2010). Do information and communication technologies (ICTs) contribute to development? *Journal of International Development, 22*(5), 625–640.
Kaawa-Mafigiri, D., & Walakira, E. (Eds.). (2017). *Child abuse and neglect in Uganda*. Springer.
Kanyam, D., Kostandini, G., & Ferreira, S. (2017). The mobile phone revolution: Have mobile phones and the internet reduced corruption in Sub-Saharan Africa? *World Development, 99*, 271–284.
Kessides, C. (2005). *The urban transition in Sub-Saharan Africa: Implications for economic growth and poverty reduction*. The World Bank. Africa Region Working Paper Series No. 97.
Kiiza, B., & Pederson, G. (2012). ICT-based market information and adoption of agricultural and seed technologies: Insights from Uganda. *Telecommunications Policy, 36*(4), 253–259.
Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica, 75*(1), 83–119.
Kosack, S., & Fung, A. (2014). Does transparency improve governance? *Annual Review of Political Science, 17*, 65–87.
Kremer, M., Duflo, E., & Dupas, P. (2011). Peer effects, teacher incentives, and the impact of tracking. *American Economic Review, 101*(5), 1739–1774.
Larochelle, C., Alwang, J., Travis, E., Barrera, V. H., & Andrade, J. M. D. (2017). Did you really get the message? Using text reminders to stimulate adoption of agricultural technologies. *Journal of Development Studies, 2017*, 1–17.
Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *The Annals of Applied Statistics, 7*(1), 295–318.
Meier, P., & Munro, R. (2010). The unprecedented role of SMS in disaster response: Learning from Haiti. *SAIS Review, 30*, 91–103.
Muto, M., & Yamano, T. (2009). The impact of mobile phone coverage expansion on market participation: Panel data evidence from Uganda. *World Development, 37*(12), 1887–1896.
Ogutu, S. O., Okello, J. J., & Oteino, D. J. (2014). Impact of information and communication technology-based market information services on smallholder farm input use and productivity: The case of Kenya. *World Development, 64*, 311–321.
Peixoto, T., & Sifry, M. L. (2017). *Civic tech in the global south*. Washington, DC: World Bank and Personal Democracy Press.
Peña-López, I. et al. (2016). *World development report 2016: Digital dividends*. The World Bank.
Sjoberg, F. M., Mellon, J., & Peixoto, T. (2017). *The effect of bureaucratic responsiveness on citizen participation*. Public Administration Review.
Thompson, M. (2008). ICT and development studies: Towards development 2.0. *Journal of International Development, 20*(6), 821–835.
The World Bank (2016). *Making politics work for development: Harnessing transparency and citizen engagement* (Conference edition). Washington, DC.: World Bank. Policy Research Report.
Young, A. (2017). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *London School of Economics*. Working Paper.