

Information Technology and Political Engagement: Mixed Evidence from Uganda

Guy Grossman, University of Pennsylvania
Macartan Humphreys, Columbia University and WZB Berlin
Gabriella Sacramone-Lutz, Columbia University

This study integrates three related field experiments to learn about how information communications technology (ICT) innovations can affect who communicates with politicians. We implemented a nationwide experiment in Uganda following a smaller-scale framed field experiment that suggested that ICTs can lead to significant “flattening”: marginalized populations used short message service (SMS) based communication at relatively higher rates compared to existing political communication channels. We find no evidence for these effects in the national experiment. Instead, participation rates are extremely low, and marginalized populations engage at especially low rates. We examine possible reasons for these differences between the more controlled and the scaled-up experiments. The evidence suggests that even when citizens have issues they want to raise, technological fixes to communication deficits can be easily undercut by structural weaknesses in political systems.

Wweak political communication channels characterize many low-income countries. Traditional aggregators of interests, such as civil society organizations, labor unions, and political parties, have limited reach, and regular high-frequency public opinion polls are all but nonexistent. Many citizens have only limited opportunities to directly communicate with politicians, usually around elections, and only very few are willing to bear the high costs of reaching out to representatives to articulate preferences outside electoral cycles. Specifically, constituents do not invest in articulating preferences if they doubt that government officials will be responsive to citizen demands.¹ Such a weak sense of political efficacy is compounded by the high cost of traditional forms of political communication (e.g., traveling large distances to meet public officials in person) that further reduce citizens’ incentive to proactively reach out to politicians in order to articulate interests, priorities, needs, and preferences.

Weak political communication channels may have important implications for the health of a country’s democratic institutions: with poor information on their constituents’ preferences and policy priorities, elected representatives have a hard time representing interests, and political parties cannot differentiate themselves in meaningful ways (Bleck and van de Walle 2013). When parties are nonprogrammatic, the accountability relationship between officeholders and voters can narrow down to local clientelistic exchange (Stokes et al. 2013). The starting point of this study is that strengthening weak political communication channels offers a promising way to begin improving political representation.

In this article, we report findings from a multiyear research project (involving three related field experiments) designed to test whether innovations in information communication technologies (ICTs) can be harnessed to improve political communication in low-income countries. Since the existence and

Guy Grossman (ggros@sas.upenn.edu) is a professor of political science at the University of Pennsylvania, Philadelphia, PA 19104. Macartan Humphreys (mh2245@columbia.edu) is a professor of political science at Columbia University, New York, NY 10025 and a research director at WZB Berlin, Berlin 10785. Gabriella Sacramone-Lutz (gs2580@columbia.edu) is a PhD candidate in political science at Columbia University, New York, NY 10025.

Support for this research was provided by National Science Foundation grant 1260631. Data and supporting materials necessary to reproduce the numerical results in the article are available in the *JOP* Dataverse (<https://dataverse.harvard.edu/dataverse/jop>). An online appendix with supplementary material is available at <https://doi.org/10.1086/708339>.

1. A low sense of (external) efficacy may be especially prevalent where governments have low capacity or low levels of legitimacy (Craig, Niemi, and Silver 1990). A sense of (internal) efficacy—i.e., the belief that one has the personal ability to participate effectively in politics (Niemi, Craig, and Mattei 1991)—can be especially weak for marginalized populations, whether defined by gender, education, wealth, or partisanship (Coleman and Davis 1976). Although voting rates are sometimes higher among the poor (Kasara and Suryanarayan 2015) and less educated (Croke et al. 2016), this often reflects differences in mobilization (or repression) in different contexts and does not extend immediately to other types of political engagement.

costs of new ICT platforms are likely correlated with features of a political system that may independently determine political engagement, assessing the effects of technological innovations on political communication is fraught with difficulties. To overcome this identification challenge, we partnered with the Parliament of Uganda and the National Democratic Institute (NDI), an international nongovernmental organization (NGO), to implement one of the largest field experiments involving political elites to date.

Our primary experiment examines a nationwide Parliament-led program that introduced a new channel for contacting elected representatives. In the terminology of Harrison and List (2004), this experiment is a natural field experiment (NFE), implemented as part of the political process in Uganda. The program established and subsidized a mobile-technology platform for political communication with the goal of increasing and diversifying citizen voice. Citizens in over 100 treatment constituencies were able to communicate with their member of Parliament (MP) by sending text messages. The design involved experimental variations in the price of messaging as well as encouragement that communicated the usage by others (“feedback”). In principle these allow us to assess how beneficial effects of communication systems depend on design, as well as help us to learn about the logic of constituents’ decision-making.

MPs representing treated constituencies could respond to messages via the platform and use the system’s functions to aggregate messages and to learn about usage patterns over time. The ICT platform was introduced to voters via twice-daily short radio ads in 19 national languages over six months. This experiment is unusual in scale—the program involved about 10 million voters—but also in nature: change in access was led by political elites and thus provided a relatively strong invitation to citizens to engage in politics.

The design allows us to examine four primary questions. First, what are the overall levels of political engagement through this ICT channel. Second, we assess whether this channel plausibly flattens citizens’ access to MPs. Third, in line with past work, we assess whether the costs of communication (price) matters—in principle flattening effects could be strengthened or weakened by the decision to send text messages through cost considerations. Fourth, we examine downstream effects, recognizing that citizens’ attitudes—especially their sense of efficacy and trust in government—could be affected by the introduction of a communication technology even if they do not elect to use it.²

2. These dimensions are derived, in part, from a simple theoretical model found in the appendix (available online), in which we explore implications arising from the fact that politicians might be both biased

The results of the nationwide field experiment are disappointing: system usage in treatment constituencies was low, and marginalized populations largely refrained from using the ICT platform. Pricing did not seem an important consideration, and evidence for downstream effects was weak at best. In fact, because of the disappointing level of citizen engagement and revealed—as compared to stated—low interest among MPs, the Ugandan Parliament ultimately decided to phase out the short message service (SMS).

Strikingly, these disappointing findings differed markedly from findings from a more controlled experiment—in the terms of Harrison and List (2004), a “framed field experiment” (FFE)—undertaken before the national program was rolled out and reported in Grossman, Humphreys, and Sacramone-Lutz (2014). Results from the FFE suggested a relatively high demand and that mobile technology could democratize political communication because marginalized constituents were willing to engage at relatively high rates and were not more price sensitive, compared to less marginalized voters. By contrast, the NFE found little citizen involvement and no improvement in differential access to political elites.

In the second part of the article we take advantage of differences between the NFE and the FFE to explore the reasons for the disappointing findings in the national experiment. Since both experiments were implemented using subjects from constituencies across Uganda, they involved similar populations, eliminating common external validity concerns—that replications tend to fail because of unobserved features of the experimental subject pool (Allcott 2015). We find that a large part of the explanation hinges on the ability of government to reach citizens to engage them in this kind of innovation. In particular our findings cast doubt on the utility of using short radio ads to elicit wide-scale participation.³ With that said, we still estimate that about 1 million Ugandans received the encouragement to contact officials. It is therefore possible that differences are due (also) to pure scale effects. We assess this possibility by looking for evidence of strategic substitution, exploiting exogenous variation in feedback on system usage.⁴ We do not, however,

against certain groups and lack information about citizens’ priorities. Although the basic logic that politicians respond more effectively to citizens when better informed is simple enough, the implications of this for what sorts of citizens are more likely to bear costs to inform politicians can be quite complex. Our model thus highlights the kinds of theoretical ambiguities that can arise in this context.

3. Plausibly, radio programming may be effective even if radio ads are not (Adena et al. 2015; Yanagizawa-Drott 2014).

4. Relatedly, see Ferrali et al. (2019), who explicitly model messaging politicians as subjected to positive externalities, which is appropriate when feedback can facilitate voter coordination (Arias et al. 2019). Substitution are also possible if free rider logics are in operation.

find evidence in support of this explanation. Instead, we find relatively strong evidence that voters doubt the efficacy of contacting their MP directly—partly counteracted through direct invitations of the form present in the FFE—and suggestive evidence that larger (structural) inequalities prevented the ICT program from having effects at scale.

This article makes several contributions to the literature on political communication, especially to our understanding of inequalities in political participation. We highlight how the underlying willingness to engage in politics—even when using low-cost impersonal communication channels—crucially depends on citizens’ belief about the effectiveness of political engagement, which itself likely depends on politicians’ response to incoming messaging.⁵ We provide below evidence, although not causally identified, that the usage of the system was tightly connected to MP’s (in)action. ICTs, we argue, in and of themselves, do not make nonresponsive politicians responsive.

The article also contributes to a growing literature on the effectiveness of using ICT innovations to improve governance outcomes (Peixoto and Sifry 2017). Past studies have focused on using ICTs to reduce absenteeism among frontline service providers (Duflo, Hanna, and Rya 2012; Grossman, Platas, and Rodden 2018), improve election integrity (Callen and Long 2014), increase engagement in local affairs (Buntaine, Daniels, and Devlin 2018), and report corruption (Blair, Littman, and Paluck 2019) and violent incidents (van der Windt and Humphreys 2016). Ours is the first study to examine the role ICTs may play in altering citizen-MP relationships in the context of low-income countries.

Our study also contributes to ongoing methodological debates on the utility of relatively small-scale controlled experiments, such as the FFE described here (and, a fortiori, “artefactual” field experiments or lab experiments), in shedding light on core political processes. Most field experiments—including many NFEs—are implemented on a small scale but seek to make claims about large-scale processes. For example, small-scale experiments may be used to test new approaches, be designed as a proof of concept, or test micrologics that arguably underlie general features of human behavior. Indeed, much of the “credibility revolution” in the study of international development is premised on the idea that small-scale field experiments can create a body of knowledge that allows promoting “what works” and eliminating programs and policies that do not (Banerjee and Duflo 2009). Yet, it is often contestable whether the results of small-scale field experiments can

accurately inform theory or form the basis for more general policy (Manski 2013). Our study distinguishes between explanations for when and why such inferences may not be valid and garners evidence for or against these different explanations.

In the remainder of this article we introduce the research questions that the different field experiments were designed to answer and present the design and results from the scaled-up national program. We then present analyses designed to assess mechanisms that could account for differences in outcomes. Our conclusions focus on the implications for efforts to democratize political communication and on the implications for learning about political processes from controlled experiments.

RESEARCH DESIGN

The field experiment we study was part of the national strategy of Uganda’s Parliament for widening citizen voice. To the best of our knowledge, it is one of the largest political field experiments ever to be undertaken with consenting political elites.⁶ Below we describe the political context that gave rise to this intervention—summarizing results from the FFE implemented before the national intervention—and describe the design of the national intervention and the data used to study it.

Political context

Uganda provides a good context for exploring changes to behavior in the wake of introducing a new political communication platform. First, Uganda shares characteristics with many low-income countries on relevant dimensions. It is in the midrange of the World Bank’s low-income economies in terms of economic development (as captured by gross domestic product per capita) and human development (as captured by Human Development Index ranking).⁷ In addition, Uganda is in the middle range of ICT ownership, use, and access among African countries (World Bank 2016). These factors strengthen confidence in the external validity of our results.

6. Our study joins a growing body of work using politicians as experimental subjects. See, e.g., LeVeck et al. (2014) and Sheffer et al. (2018) on politicians’ decision-making and Grossman and Michelitch (2018) for politicians’ response to disseminating information on their performance in office. Such studies can raise ethical concerns insofar as they involve interventions in democratic processes. In this case, however, we highlight that the initiative was with the explicit consent of politicians and was owned by the Parliament of Uganda, operating through the Parliament’s website, with the intention of strengthening democratic processes. The uSpeak program had no deception of any form.

7. Low human development countries are ranked between 148 (Swaziland) and 188 (Central African Republic). Uganda is ranked 163 (in 2016).

5. For recent studies making similar claims, see Christensen and Ejdemyr (2020), Grossman, Michelitch, and Santamaria-Monturiol (2017), and Sjoberg et al. (2017).

Second, data from Uganda support the assumption of weak political communication channels leading to a dearth of information in the hands of politicians. Consider results culled from a survey the research team conducted with Ugandan MPs at baseline. We find that the majority of surveyed MPs describe themselves as feeling insufficiently informed when they vote in plenary and in committee meetings. In other work, surveyed Ugandans report that elected politicians do not frequently elicit voter opinions (Grossman et al. 2017). This evidence suggests that the context is one in which there is an unmet demand for greater information.

Third, results from the FFE conducted before the launch of the national field experiment further point to Uganda as a good context for studying the questions at hand. Specifically, findings from the FFE suggest not only that there is an underlying demand in Uganda for contacting one's MP via a text-messaging platform but also that IT communications do not necessarily widen the participation gap between more and less marginalized populations. We briefly describe the FFE below.⁸

The FFE, undertaken in 2011, was delivered alongside a survey conducted in every parliamentary constituency in Uganda using a nationally representative sample. The FFE sought to assess whether demand existed and to explore the validity of the concern that IT-based communication platforms exacerbate existing inequalities in political access. At the end of the survey, sampled respondents were invited to send a text message to their MP at randomly assigned prices. Discussed in more detail in Grossman et al. (2014), the communication recorded in the FFE suggests that a sizable number of citizens (about 5% of the study sample) value the opportunity to contact their MPs via SMS.

In addition, usage rates in the FFE were no lower among more marginalized populations, possibly reflecting the fact that these populations have fewer opportunities to access politicians and therefore place a higher value on impersonal and inexpensive ICT channels. Experimentally manipulating the price of sending a text message to one's MP, we further found, as expected, that reducing the cost of communication encouraged usage.⁹ Moreover, consistent with the idea that marginalized populations place a higher value on cheap impersonal communication, we found that marginalized populations were not more sensitive to the cost of political communication than less marginalized populations.

8. A more detailed description of the FFE can be found in Grossman et al. (2014).

9. Usage was almost 50% higher for those randomly assigned to a free SMS treatment arm, as compared to those assigned to a treatment group that was not offered any subsidy for texting their MP.

The FFE confirmed that Uganda offers a good context to examine the implications of harnessing technological innovations to improve political communication and that ICT platforms have a potential to alter citizen-MP relations and "flatten" political access. However, the setup of the FFE also had some limitations. For example, it allowed only a "one-shot" opportunity to communicate with MPs and thus was unable to examine usage patterns over time, in which citizens' behavior is (also) a function of both the usage of other citizens and the response of their MP to past messages. Moreover, it was implemented in the context of an in-person survey in which subjects interacted with enumerators regarding their political views. While such personal interaction ensures that subjects are aware of the program, it is also prohibitively costly. Thus, there is no guarantee that using mass communication channels at scale (such as radio) would result in a strong first stage (i.e., wide-range program awareness). The personal interaction with enumerators may have also made politics more salient to interviewed subjects, further strengthening the invitation to use the platform. The personalized invitation to contact one's MP may have also increased both the sense of empowerment and civic obligation to raise one's voice. It is also possible that subjects perceived the FFE as closer to a civil society effort than an official government program.

These considerations raise the question of whether similar effects would be found when the ICT service was brought to scale and shifted from being a researcher-led initiative to being an institutionalized part of national politics. The field experiment described in the next section was designed to address these concerns.

The uSpeak initiative

As part of the Ugandan Parliament's national strategy, a case management platform hosted in the National Parliament was developed, allowing citizens to send messages to their MP via SMS or a voice call to a call center. MPs randomly assigned to participate in the program (uSpeak) were given access to the platform and trained in its use. The platform allowed MPs to log onto a dashboard where they could read tagged SMS messages from constituents, reply, and see simple descriptive statistics about the messages they received, such as what the priority issues in their constituency were within a selected time frame. A screen shot of the query dashboard is presented in appendix figure 2. Only treated MPs were able to receive messages from their constituents via the case management system.

The ICT platform was promoted to citizens through 30-second radio advertisement spots, played twice daily on local radio stations over the study's six-month period. The radio ads were in local languages and featured a skit in which actors

portraying constituents talked about how uSpeak could be used to draw the MPs' attention to important issues, specifically service delivery deficiencies. These skits were first tested using focus groups. A second tier of randomly assigned treatments—price and feedback—was also delivered via the radio ads.

Treatment 1: Elite participation. The NFE involved 186 MPs who volunteered to be part of a six-month pilot. It was expected that, if deemed successful, all MPs would be phased into the program at the end of the study. Given the sensitivities of providing a new service to only some constituencies, it was agreed that MPs would be selected into the program using a public lottery managed by NDI. Block randomization was used to assign MPs to treatment groups; MPs were sorted into bins on the basis of their type (woman MP or constituency MP), party, and region.¹⁰

Treatment 2: Variation in price. To assess the effects of price on usage, Parliament randomly varied the cost of sending a message to MPs via the uSpeak system, across and within constituencies. Each constituency was assigned three months in which uSpeak would be provided free of charge and three months without any subsidization. Being sensitive to potential sequence effects, all possible sequences of full price and free months were randomly assigned to constituencies in the treatment group using a blocked design. Note that while the variation in prices in the first period provides a clean separation into price groups, for identification on the basis of variation in subsequent months we must assume no carryover effects.

Treatment 3: Variation in feedback. In order to examine whether information on others' usage encourages greater system usage, we added a feedback treatment arm delivered through modification of the base radio ads. In one version, voters heard that others had been sending messages to the system about the need to do more in the educational sector. A second variation also highlighted the educational sector but without communicating that others had been using the system to lobby in that area. To the extent that there are complementarities in public goods messaging, we expect that hearing that others are sending messages about education should

increase the willingness to contact one's MP. Indeed, our feedback skit was written explicitly in a way that made this sort of complementarity more apparent to radio listeners.¹¹ We make use of this variation to help unpack reasons for differences in results between studies.

We selected eight unique price sequences and six unique combinations of the feedback treatment that together produced 48 unique combinations of price and feedback sequences. These were assigned in a balanced way to treatment constituencies, resulting in roughly two constituencies of each unique treatment schedule. In appendix figure 5 we provide an example of treatment schemes for a subset of constituencies.

Data

Data for testing the effects of the uSpeak program come from four sources: (1) a baseline survey of Ugandan adults randomly drawn from all constituencies in Uganda, conducted immediately following the 2011 Parliamentary election; (2) the SMS messages sent by constituents to the uSpeak system, tagged with the date and time they were received; (3) a callback phone survey we conducted with uSpeak users; and (4) an endline survey of a nationally representative sample of Ugandan adults in a subset of the uSpeak constituencies. In addition, as described below, we conducted a follow-up experiment with about 3,000 Arua district residents to help adjudicate some of the conflicting findings between the NFE and the FFE.

MAIN RESULTS

We focus on core results related to overall usage, flattening (the characteristics of participating populations), price effects, and downstream effects. We note that usage and flattening are not experimental treatment effects in the usual sense; rather, they are levels assessed under controlled conditions. Price and feedback effects draw on randomized variation within treatment and downstream effects draw on randomized MP participation in the intervention, as described above. Analyses implemented to explain our findings on usage are described in the discussion section.

Low rates of communication

Unlike the FFE described above, system usage in the NFE was very low. Despite twice daily radio ads and price subsidization throughout the country, MPs in the treatment group received a total of 1,946 messages during the six-month study period.

10. Each bin was used to implement a separate public lottery, with a target number of MPs selected into treatment on the basis of that MP type's prevalence in the subject pool. Block randomization was used not simply to improve balance in expectation but also to improve ex post equality between parties in participation. See app. sec. 3.5 for additional information on the public lottery.

11. By contrast, if people view text messages as substitutes, then hearing that others are using the IT system could exacerbate the collective action problem. The feedback ads are shown verbatim in app. sec. 3.3.

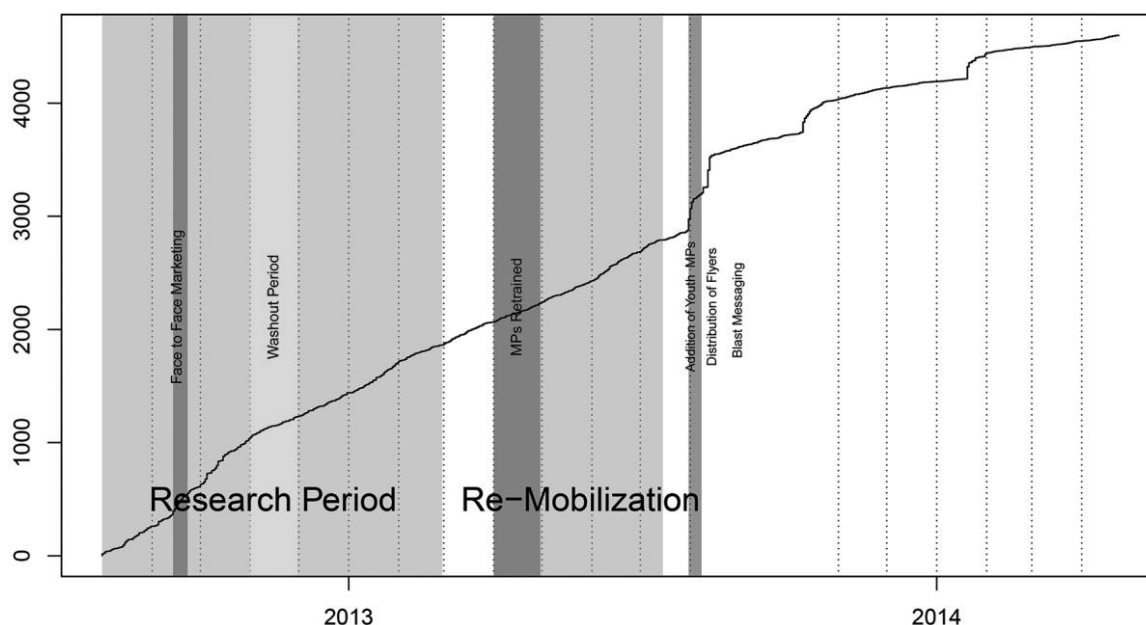


Figure 1. uSpeak natural field experiment: cumulative messaging over time. Light-gray area represents the washout period in which no radio spots were played. Medium-gray areas denote the period with experimental variation. Usage in the postexperimental period includes attempts by Parliament and the National Democratic Institute to further encourage usage. Color version available as an online enhancement.

Using the most recent 2014 population census, we estimate conservatively that the radio ads were played over an area where 10 million voters reside. This thus corresponds to a monthly usage of about 1 in 30,000. Figure 1 shows the cumulative messaging over time, extending beyond the study period to show usage in the poststudy period including various periods in which an assortment of mobilization efforts were used by Parliament and NDI—none of which produced sustained effects.

A broad categorization of the types of messages suggests that, as with the FFE (reported in Grossman et al. 2014), a large share of messages were for local public goods or local community interests with a much smaller set for national or policy concerns. A much larger share of messages here were of a more personal nature, accounting for nearly half of messages sent compared to at most 10% in the FFE (see app. table 2).

No flattening effects

One of the key findings of the FFE was that the share of marginalized populations—such as women and the poor—among system users was higher than the share of marginalized constituents participating in traditional forms of political engagement. That finding formed the basis of our conclusion that ICT platforms have a genuine potential to flatten political access.

To assess flattening in the national experiment, we conducted a phone survey of system users. Using a call center that the research team had set up, local enumerators con-

tacted all uSpeak users no longer than two months after they had sent a text message to their MP. The short callback survey was designed to elicit information on users’ demographics, whether they received a response from their MP, and general satisfaction with the ICT service.

Comparing results from our callback survey to information culled from the FFE, it is clear that the scaled-up national program failed to replicate the flattening effect identified in the FFE. Specifically, the users of the uSpeak system were wealthier, more highly educated, and overwhelmingly male, compared to those sending text messages in the FFE. Put plainly, the uSpeak program failed to elicit participation from marginalized populations in the way political actors expected. Figure 2 provides information on the distribution of wealth, gender, and education, across the two field experiments.

The patterns here suggest that usage is relatively higher among wealthier, more educated, male citizens.¹² However, the substantive importance of this is lessened by the fact that usage was exceedingly low overall. Had there been substantial flattening, this would have occurred only within a likely small part of a politician’s information set and thus been very unlikely to have had any effect on politicians’ subsequent behavior.

12. Note that flattening is a summary of heterogeneous usage. Overall low usage could (in theory) reflect a moderate usage among marginalized citizens combined with no response among the nonpoor. Thus, testing for flattening is analytically distinct from exploring overall usage.

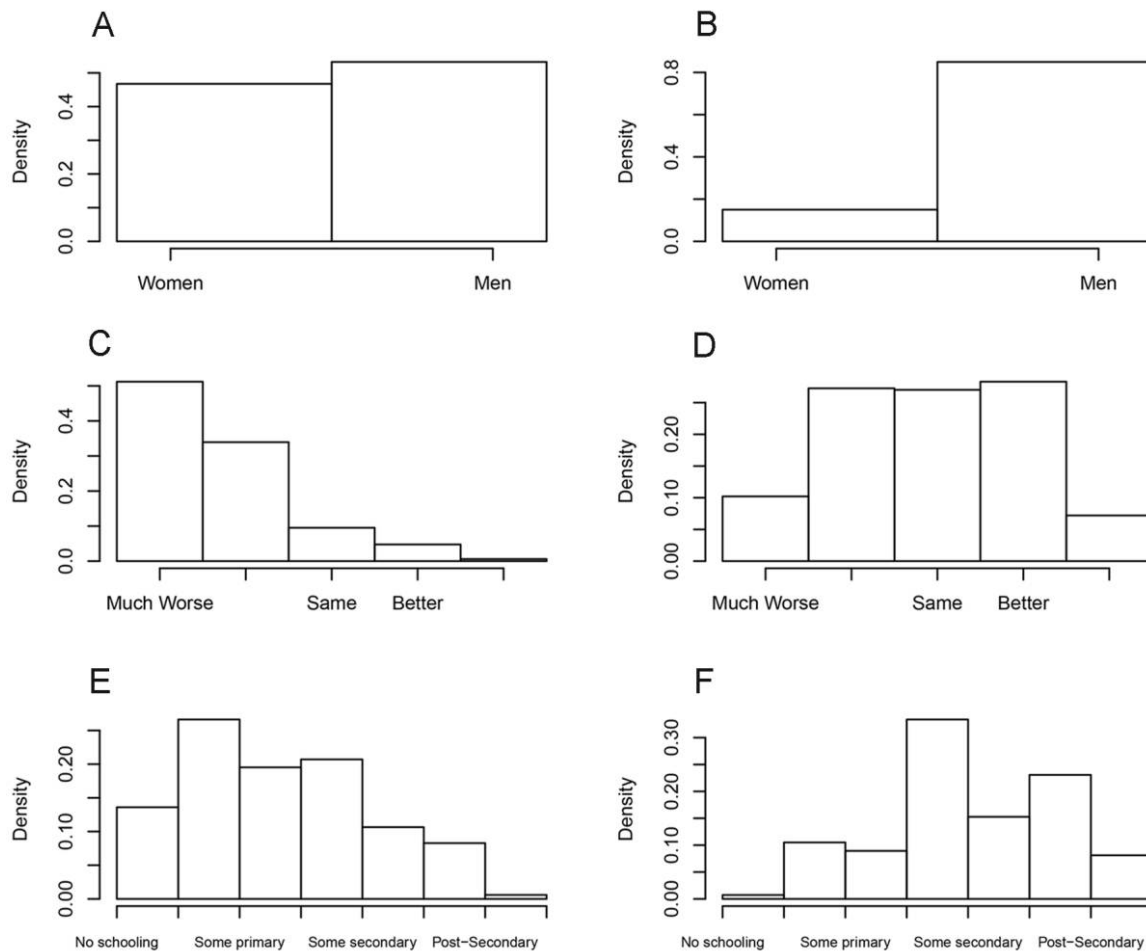


Figure 2. Demographic differences: users in the framed field experiment (FFE) compared to users in the natural field experiment (NFE), uSpeak. A, FFE gender; B, NFE gender; C, FFE wealth; D, NFE wealth; E, FFE education; F, NFE education. Users in the scaled-up NFE were more likely to be male, better educated, and wealthier than users in the FFE. Data source: phone surveys of all system users.

Insensitivity to price

Unlike the FFE, we find no evidence of overall sensitivity to price in the scaled-up national program.¹³ Monthly rates of messaging in the free and full-price treatment conditions are in fact almost indistinguishable. Testing for a price effect more formally, we run a linear regression of the number of messages received in a given month on price—a binary variable that takes the value 1 for full price and 0 for months of free messaging—controlling for the month of feedback treatment indicator and MPs fixed effects. Results presented in table 1 suggest that contrary to the FFE, in the scaled-up national program, price did not significantly affect usage. The substantive magnitude of the estimated effect is tiny, as is the upper bound of the confidence interval—at the upper bound

13. Recall, the price effect is an experimental property; monthly prices are randomized over the study population irrespective of actual usage.

switching to a free system is associated with less than one additional message per constituency per month (compared to a 2 percentage point increase in messaging from the FFE; a rate that would have resulted in about 30 additional monthly messages per constituency).¹⁴

No evidence of downstream effects

Thus far we have shown that system usage in the scaled-up uSpeak program was low and that fully subsidizing the cost of messaging did not increase voters’ proclivity to contact their MP via SMS. Notwithstanding the low rates of usage, it is possible that uSpeak had a positive effect on voters’

14. The divergence observed in price effect across experiments is reminiscent of the way subjects of controlled laboratory experiments react to even small monetary manipulations that are inconsequential outside laboratory settings.

Table 1. Usage as a Function of Price and Feedback

	(1)	(2)
Price	-.081 (.262)	-.079 (.262)
Education prompt		-.275 (.453)
Education plus feedback prompt		-.132 (.452)
R ²	.055	.055
Adjusted R ²	-.145	-.149
F-statistic	5.251*** (df = 6; 544)	3.972*** (df = 8; 542)

Note. Dependent variable = messages per month. Standard errors in parentheses. *N* = 660.

* *p* < .1.
 ** *p* < .05.
 *** *p* < .01.

sense of efficacy and their satisfaction with politics in Uganda. This would be the case if citizens view the existence of the ICT platforms, irrespective of one’s own usage, as an important tool for strengthening citizen voice. This was a goal of the intervention, and we report on it here briefly. Results in this section use experimental estimates of the effects of the intervention, exploiting the random assignment of the scaled-up program.

To test for the effect of the national program on voters’ efficacy, we turn to our endline survey. The survey, which took place in July–August 2014, included 2,714 adult respondents from 76 constituencies and 304 villages in 52 districts across Uganda.¹⁵

To measure efficacy, we asked survey respondents whether they agree with the following statement: “People like you can do things that can have an influence on the actions of . . . [your constituency MP].” We then repeated the question for the president, district chair, and traditional leaders, which we use as placebo tests (rather than multiple tests on a common hypothesis). Our key dependent variable is a binary indicator that is equal to 1 for the 60% of respondents who had agreed that citizen action could influence their MP. We then run a simple ordinary least squares model regressing the efficacy outcome on a treatment indicator and district fixed effects.

Figure 3 shows that we did not find statistically significant evidence of downstream effects on support. The magnitude of the estimated effect is, however, reasonably large and is per-

haps the most optimistic finding from the study. Yet probing further does not increase confidence in these results. In the graph we also report results from four placebo tests, assessing increased confidence in leaders who are not related to uSpeak. Overall the estimated effects are also large for these results and indeed significant in two cases. While surprising, the pattern suggests that the intervention did not increase the efficacy of citizens with respect to MP engagement relative to the effect on engagement with other political actors.

DISCUSSION

Experimental findings from the national program conflict with the results from the FFE. Notably, uSpeak resulted in low usage, even when the service was offered to voters at no cost. Moreover, confirming concerns that ICTs would exacerbate existing inequalities in political access, when uSpeak was used, it was by and large by citizens whose voice is already more likely to be heard. In other words, the groups that have the weakest access to political processes were also the least likely to use the new ICT platform.

Here we explore some of the reasons that may account for the low usage of the uSpeak system. In appendix section 8 we also assess several explanations for the fact that—contrary to the FFE—marginalized populations were significantly less likely to use the new ICT platform. This question is of somewhat less importance because of the low usage of the uSpeak system—with low take-up, any potential flattening is only with respect to a less important information source, as mentioned above.¹⁶ For similar reasons, we do not explore the rationale behind the lack of downstream effects, since the fact that the scaled-up national program generated such weak first-stage results makes it less surprising that voters and politicians’ attitudes and behavior were not affected by the introduction of uSpeak.

Our goal in closely examining the causes of the low usage in the NFE is not merely to account for these diverging results but rather to use the analysis to derive substantive insights regarding the role ICTs can currently play in improving political communication in low-income countries. Although the FFE led by the research team was meant to capture the key features of the scaled-up national ICT platform, the introduction of relatively tight experimental controls introduces a number of differences.

We first explore whether differences in results could be due simply to differences in samples, exploiting the fact that the NFE constituencies are a subset of the FFE constituencies. We next explore the explanatory power of two external features of

15. See app. table 3 for descriptive statistics of those survey respondents.

16. We thank a reviewer for pointing out the implications of low overall usage on the political salience of results on flattening.

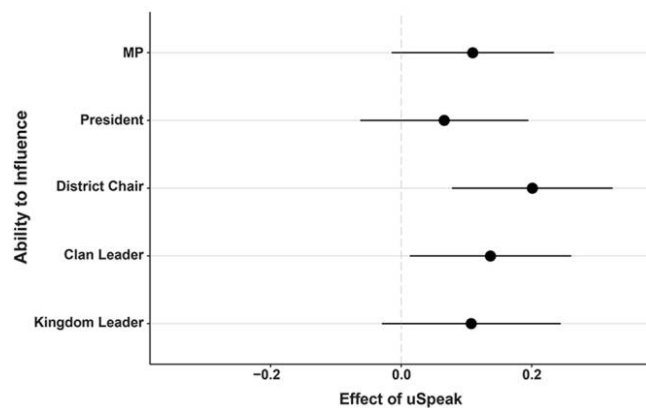


Figure 3. Marginal effect of uSpeak on political efficacy measured as respondents' perception of their ability to affect their member of Parliament.

the NFE—which are common to interventions that are scaled up from controlled pilots to larger programs—that may have been consequential. We refer to these as “scale” and “agent” effects. In addition, we examine the implications of subtle differences in the delivery of the treatment. These design effects may be especially relevant for interventions that involve the dissemination of information to subjects.

Changes in scale are often described as a problem of general equilibrium effects (Deaton 2010). This concern is of particular salience when treatment effects are sensitive to the share of those treated in the population. Scale effects are of special concern when subjects can accurately infer the magnitude of a program from its delivery method, as is clearly the case in our study. In our setting, it is quite possible that collective action problems get altered substantially as scale increases. Insofar as political communications complement each other, or substitute for each other, increases in scale could lead to greater or lower overall levels of communication.

A third possible reason for the low usage relates to agents. Whereas the research team implemented the FFE, the Parliament of Uganda and NDI led the scaled-up national program.¹⁷ In our case, this change in agents might have affected citizen expectations regarding the responsiveness to their messages. In other words, the fact that the scaled-up national intervention was implemented by Parliament rather than by researchers may have reduced the incentives of the target population to engage.

The fourth possibility relates to experimental design and, specifically, to the possibility that details of the mode of treatment delivery—the nuts and bolts of executing field experi-

ments—mattered a great deal for citizens in deciding whether to communicate with elites. We focus here on two possibilities. The first is that the method of delivery (radio spots) introduced a treatment compliance effect: that Ugandans were simply unlikely to hear or internalize appeals issued through mass media, and not less likely to respond, conditional on hearing. This mode of delivery differs from the FFE, where the enumerators ensured that respondents unambiguously heard and internalized the information on the SMS platform.

Closely related is the possibility that different methods of delivery have varying degrees of an (implicit) invitational effect. It is possible that communication was relatively high in the FFE, not simply because the in-person survey context ensured awareness of the new ICT platform but also because the enumerators had personally invited respondents to contact their MP. Communicated in the context of a survey, such invitations may appear as more personal encouragement to engage in politics. A direct personal invitation has an empowering effect, signaling receptiveness and the possibility that political communication will make a difference. These last two design mechanisms are closely related yet distinct: one is about whether an invitation was empowering and deemed personal, and the other is whether an invitation gets heard at all.

We use a number of strategies to adjudicate between these five explanations. First, we exploit a feature of the scaled-up field experiment in which there was variation in the feedback provided to voters on the behavior of others. This allows us to examine whether exacerbating collective action problems due to scale can, at least partially, account for the significantly lower usage. Second, we conducted a citizen endline survey with a nationally representative sample, which allows us to assess—albeit with some lag—ex post differences in treatment compliance. Third, we hired a Ugandan private marketing firm to examine whether the radio stations NDI had contracted indeed played the ads according to the experimental design. Fourth, we implemented an additional “mechanism experiment” in one district in Uganda in which we specifically varied the invitational component. Table 2 summarizes the list of potential explanations and the source that was used to explore their explanatory power.

Randomization bias

One concern is that different results arise from differences in the study population and the target population.¹⁸ This is a form of what Heckman (1991) has called “randomization bias” in the sense, here, that control of the research team

17. Differences in agents across scales are common: e.g., the Millennium Villages initiative sought to assess the scope for government-led development change by examining an intervention in which the government was not a primary actor.

18. We thank a reviewer for highlighting this point of difference.

Table 2. Differences between the FFE and the NFE

Type	Data Source
Sample differences (randomization bias)	FFE data (Grossman et al. 2014)
Scale effects	National experiment (feedback treatment)
Agent effects	Users survey data (callback)
Design effects:	
Treatment compliance	Citizens survey data (national sample), radio monitoring data
Invitational effects	Follow-up mechanism experiment

Note. FFE = framed field experiment; NFE = natural field experiment.

over experimentation (or sample) selection ensured that all constituencies took part in the FFE whereas only one-third of constituencies took part in the NFE.

Selection into experimentation can mean that researchers do not have information on how nonsubjects would perform in target contexts of interest. In our study, the selection took place for the scaled-up experiment (the target) and not in the controlled experiment. This is different from the usual situation in which the bias is in the controlled setting rather than in the target case. It means, however, that we are in a position to assess to some extent how great randomization bias in the scale-up is. To explore this, we return to data on the three indexes from the FFE: access, political engagement, and marginalization (see table 3). For each of these we compare mean values for the (self-selected) NFE sample, the uSpeak constituencies (randomly sampled from the NFE constituencies), and the endline constituencies (randomly sampled from the NFE constituencies) against the FFE sample. None of these differences are large. Moreover, on these key features there is no significant difference between the FFE and NFE constituencies. Most importantly the rates of sending messages in the FFE were essentially identical, and so we conclude that this difference in samples is unlikely to explain differences in behavior in the NFE.

Scale effects

Political communication is subject to collective action problems. If many others are visibly lobbying a politician, free ridership may become more likely if messages work as substitutes. By contrast, visibility of lobbying can improve voters' coordination leading to a cascade of usage, assuming a sufficient number of voters view messages as complements, as in Ferrali et al. (2019).

We can imagine two ways in which a logic of this form can play out. First, if politicians are more informed about others, and better able to target resources to them, this can increase incentives to provide a politician with information on one's own preferences. Second, scale may result in lower individual contributions through a simple logic of substitution for members of a given group. In the extreme case in which information from constituency members were perfect substitutes and citizens faced linear costs, increases in potential information providers would not alter the amount of information provided, in equilibrium, which in turn implies a corresponding reduction in per capita information provision.

Conversely, one could construct examples in which when many others are lobbying for a common good there may be increasing returns to lobbying (or here too, there may also be increasing incentives to free ride). In short, if the incentives to use technological innovations for political communication depend on the perception of how others are engaging (Ferrali et al. 2019), then outcomes at a small scale may look very different from outcomes at a large scale. From this perspective, weaker participation from the scaled-up program may reflect a simple failure of collective action.

In order to understand whether changes in scale induced free riding, we look for differences in outcomes due to our feedback treatment. Recall that in the scaled-up national experiment, we exogenously varied the information constituents received about the level of activity by other voters in previous periods. In particular, a random subset of constituencies was informed, through the short radio ads, that other voters had been using the system to mainly raise issues around education. Under a free riding logic, such information would depress engagement among those exposed to it.

Returning to table 1, however, we find no evidence of sensitivity of engagement to information on usage by others—neither the difference between feedback on education messages and standard marketing spots nor differences between

Table 3. Mean Indexes across Experimental Samples

Index	FFE	NFE	uSpeak	Endline
Access	0	.017	.020	.086
Political engagement	0	-.040	-.056	-.018
Marginalization	.500	.500	.503	.495
Sent SMS	.023	.021	.024	.023
N	7,582	3,651	2,023	1,567

Note. Sample averages and out-of-sample means. FFE = framed field experiment; NFE = natural field experiment; SMS = short message service (text). Access and political engagement indexes are standardized (mean equals 0 and standard deviation equals 1). Marginalization index is based on percentile.

education marketing with and without feedback is significant. This is consistent with a set of analyses we conducted on the data from the FFE in which we found no evidence for strategic engagement with the system. We conclude that scale, by itself, does not seem to be a key factor driving our divergent results in terms of overall citizen communication.

Agent effects

Another possible reason for the low system usage witnessed in the scaled-up national program is agent effects. Citizens' usage of mobile messaging plausibly increases with the belief that there is a receptive representative at the other side of the interaction (Sjoberg, Mellon, and Peixoto 2017). Citizens will be less likely to communicate if they think that a political actor cares less about their interests. More subtly, they will also be less likely to participate if political actors are better informed about the interests of rival constituents.

Which system should voters expect to produce greater responsiveness by politicians? On one hand, unlike our FFE, the scaled-up national program is formally owned and led by Parliament, which signals some level of commitment by politicians. In addition, the dynamic nature of the scaled-up program (i.e., the ability of MPs to interact with citizens directly via the ICT platform) further allows MPs to signal their responsiveness directly. This sort of dynamic reciprocal relationship could not have been established in the one-shot controlled FFE. On the other hand, in the scaled-up program, the communication between citizens and politicians was direct, whereas in the FFE this relationship was mediated by the research team, which was responsible for delivering the messages to survey respondents' respective MPs. Citizens may believe that their MP will take their messages more seriously if researchers or an NGO mediates the relationship between voters and representatives, for example, if it follows up in case some messages get ignored. Thus, it is hard to predict a priori how the change in the implementer's identity would affect citizen communication. We explore (nonexperimentally) agent effects in two ways.

First, we use the callback survey ($N = 2,517$ uSpeak users) to calculate an MP's response rate at the constituency level and then test for a correlation between the MP's responsiveness and the volume of messaging at the constituency level. We find that only 9% of uSpeak users report ever hearing back from their MP; in fact, in almost half of the treated constituencies (44) not a single uSpeak user had received any response from his or her MP. Moreover, analyzing system login information, we find that the majority of MPs did not read many (or any) of the messages sent to them. As expected, we find a positive correlation between messaging and responsiveness, which is

consistent with citizens' low engagement being a rational response to their MP's (in)action during the scaled-up study period.

Second, although the callback survey analysis focuses on system users—a self-selected group—we, nonetheless, can also assess whether broader expectations regarding MP inaction may have contributed to the low usage rate among the general population. Here we examine responses in the citizen endline survey, when our nationally representative sample was asked to indicate reasons for why people might not use SMS platforms such as uSpeak to communicate with their MPs. Appendix figure 8 (*bottom left*) provides information on the share of respondents in treatment constituencies who indicate each possible reason. Tellingly, we find that close to 50% of respondents report that they would not send a message because they do not expect their MP to be responsive, and about a quarter report a reluctance to contact their MP via text messaging out of fear of bad repercussions.

We do not have information on the expectations of responsiveness from MPs in the (one-shot) FFE and so cannot compare those data directly. Nevertheless, the statements by citizens and the very weak responsiveness by politicians suggests that the low engagement with the scaled-up program was a rational response on the part of citizens.

Treatment compliance

It is possible that the difference between the NFE and the FFE is simply due to an insufficiently strong first stage; that is, the radio ads had a limited reach, and thus an overwhelming majority of constituents never heard of the uSpeak program. To test for compliance effects we asked respondents in our citizen endline survey directly whether they have ever heard about uSpeak. The survey was implemented in 50 parliamentary constituencies approximately one year after the six-month radio campaign, although at a time when the uSpeak system was still active and promoted by Parliament. In light of the time gap, we used a deliberately strong prime, which entailed playing the original radio ad and asking respondents whether they had heard of the service (uSpeak) the ad sought to promote.

Starting with the raw data, we find that about 17% of respondents in control constituencies and 24% of respondents in treatment areas self-report that they ever heard of uSpeak.¹⁹ Note that control respondents are not necessarily misrepresenting their knowledge of the program; this is because radio signals normally have a range that encompasses more than a

19. We note that the reported rate of compliance likely is an upper bound due to the possibility of social desirability bias.

Table 4. First Stage for Constituency Members of Parliament Only

	(1)	(2)	(3)	(4)
Treatment	.125*** (.022)	.120*** (.022)	.105*** (.032)	.095*** (.034)
Education		.026 (.038)		.058 (.043)
Income		-.011 (.033)		-.036 (.035)
Constant	.089*** (.013)	.080 (.067)		
Block fixed effects	No	No	Yes	Yes

Note. Dependent variable = knowledge of uSpeak. Standard errors in parentheses. $N = 50$.

* $p < .1$.
 ** $p < .05$.
 *** $p < .01$.

single parliamentary constituency.²⁰ Testing for a first stage more formally, we take a conservative approach: first, we code respondents who live in constituency j as treated if either their constituency (usually male) MP or their district (female) MP was assigned to the uSpeak program, and then we calculate the share of the constituency hearing about uSpeak, by adjusting survey sample weights. Regression results at the constituency level, using inverse propensity weights based on both constituency and district assignment probabilities, are reported in table 4. We find a large, positive, and significant first stage (col. 1). This result is robust to whether we control for (aggregated) individual-level covariates (col. 2) and whether we add fixed effects for the randomization stratification blocks (cols. 3 and 4).

Second, we estimate constituency-level SMS sending rates (take-up) as a function of hearing about uSpeak, using two-stage least squares regressions. This analysis allows us to compare directly the take-up rate conditional on hearing about the program in the NFE against that of the FFE; it also serves as a reality check for the first-stage regression since the dependent variable here is a behavioral measure derived from the uSpeak database. As reported in table 5, hearing about uSpeak has a take-up rate of 0.002 with stan-

20. Hearing the message in control areas does not imply noncompliance since the ads were tailored to employ the name of treated MPs only. Control subjects could be aware that others were treated, but this does not make it possible for them to take up the treatment.

dard error of 0.001; in other words, the take-up rate is between 1/4 and 1/10 of 1%. Comparing this to a take-up rate of about 5% in the FFE (which is larger by a factor of about 25), we can confidently reject the null of no difference between NFE take-up among compliers and our estimated FFE take-up.

This analysis has two implications. On the one hand, the effectiveness of radio as a marketing device is not strong. Indeed, when probing deeper about respondents' knowledge of the uSpeak program we find that only 6% of treatment respondents were able to confidently say that their MP had participated in the program. Moreover, when asked to repeat the four-digit short code, less than 0.5% of treated constituencies claimed to know the short code to send a text message to their MP, and an additional 3% reported that they once knew the number but have since forgotten it. These findings strongly point to the limitation of radio marketing to garner sufficient awareness of the new service.

On the other hand, this is clearly not the full story. Take-up differential cannot be fully explained simply as a function of a weaker first stage in the NFE since the two-stage analysis suggests that the effects on those who do get the message are much weaker than in the FFE.²¹

Invitational effects

We turn to explore the possibility of a second design effect, namely, that the marketing tools used to inform citizens about a new service or program likely have (unintentional) invitational effects. Recall that the two experiments differed in their mode of marketing: whereas the scaled-up national program used 30-second radio ads, in the FFE, respondents were invited by enumerators to contact their MP in the context of an in-person survey. As mentioned, direct personal invitation may have an empowering effect, or it may signal greater government responsiveness. Multiple logics could underpin these effects: an invitation could in principle change voters' beliefs about how much the politician cares about their welfare, as well as about the politician's knowledge of citizens' preferences. As further discussed below, if such invitational effects operate differently for marginalized and nonmarginalized populations, this could account for the differences in observed flattening effects.

To assess the role personal invitations play in the decision to politically engage using an ICT platform, we implemented a third (mechanism) experiment. To do so, we used an existing SMS platform, UBridge, which has been operating in the Arua

21. A back-of-the-envelope calculation suggests that if messaging got through to 10% of up to 10 million subjects and had 5% of these responded, there would have been 50,000 messages entering the system.

Table 5. Second Stage for Constituency Members of Parliament Only

	(1)	(2)
Knowledge of uSpeak	.002*** (.001)	.002* (.001)
Constant	-.00016** (0)	-.00027 (0)
Block fixed effects	No	Yes

Note. Dependent variable = uptake. Standard errors in parentheses. $N = 50$.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

district since late 2014. UBridge was developed in partnership between UNICEF's Ureport platform and Uganda's Governance, Accountability, Participation and Performance project.²² Unlike uSpeak, which connects citizens with national MPs, UBridge was designed to open a new channel of communication from citizens to local government officials to specifically report problems of public service delivery. UBridge was launched as a pilot study in over 100 villages across Arua. A study evaluating the effect of getting access to the UBridge system is underway and is not the subject of this article. At the time of our mechanism experiment, UBridge had 4,568 registered users, out of which 2,720 were explicitly verified by the research team.²³

On June 13, 2015, UBridge conducted a baseline poll using a robocall system asking users about their attitudes toward budgetary processes. The key outcome of interest is a binary variable that receives the value 1 if the UBridge user responded to the poll and 0 otherwise. Of the 2,720 verified users, 12% responded to the opinion poll and shared their views with UBridge. To explore the role of direct invitations on levels of ICT-based political engagement, we asked UBridge to run a modified version of its baseline poll by experimentally introducing a modest variation in their outreach activities. All users would be invited to participate in an opinion poll regarding taxation, similar to the previous UBridge poll. In a randomly selected treatment group, however, UBridge preceded the call with a set of (blast) text messages that explicitly invited participants to take part in the weekend poll and that high-

22. With some loss in external validity, our design aims to keep the treatment compliance effect constant by focusing on respondents in the UBridge system. We hope that parsing the outcome compliance effect will be the focus of future studies.

23. We verified the identity of registered users through a call center that we set up with the help of Innovation for Poverty Action, Uganda.

lighted the importance of individual responses. Further details on the block randomization used in this experiment, as well as the full text of the treatment text messages, are provided in appendix section 7.

Our primary measure is the response (or nonresponse) by UBridge users to the weekend opinion poll. The encouragement text messages were delivered on June 24, 25, and 26, 2015, and the poll took place on June 26. We estimate average treatment effects using a regression that accounts for block fixed effects. Our analysis takes account of the variables used for blocking but introduces no further controls. Our primary regression uses only the verified subset of UBridge users, whereas our secondary analysis includes all registered users, whether or not they have been positively verified.

Results, reported in table 6 (col. 1), suggest that invitation had a large positive effect on response rate: 2 percentage points from a base rate of 9.4% for the control group (although it did not have a differential effect by gender: cols. 2 and 3). These results are consistent with findings reported by Grossman et al. (2017) in a similar context and by Dale and Strauss (2009) and Malhotra et al. (2011) in the United States. We note that even though the invitation tested in the mechanism experiment was relatively weak (three text messages) compared to the in-person invitation used in the FFE, it was able to increase participation rates by over 20%. The evidence at hand allows us to conclude that, consistent with insights from the voter mobilization literature (e.g., Green and Gerber 2008), more personal invitations can have a powerful effect on rates of participation. We can further conclude that, at least in low-income countries with characteristics similar to Uganda, short radio ads likely represent a marketing strategy that is too impersonal to mobilize large-scale participation.

Table 6. Mechanism Experiment

	Base (1)	Primary (2)	Secondary (3)
Invitation	.021* (.011)	.021* (.011)	.019** (.008)
Flattening (male × invitation)		.0004 (.024)	-.003 (.017)
R^2	.153	.153	.165
Adjusted R^2	.108	.107	.115
N	2,717	2,717	3,957

Note. Dependent variable = provided policy input. Standard errors in parentheses. "Male" normalized to have 0 mean.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

CONCLUSION

This study integrates three related field experiments designed to assess whether innovations in ICTs can be harnessed to improve weak political communication, prevalent in many low-income countries. Evidence from an FFE conducted before rolling out a national program suggested not only that there is underlying demand to contact representatives using mobile technology but also that ICTs have a genuine potential to increase levels of political engagement in a way that flattens access for marginalized populations. By contrast, when brought to scale using an NFE implemented nationwide, there are significantly lower levels of citizen engagement, with marginalized populations especially refraining from using the ICT platform to raise their voices. These results have implications for theory, policy, and research methodology.

Our study contributes primarily to our understanding of the promises and pitfalls of ICT-based political communications, at least in the context of low-income countries. Consider four findings that help account for the weak usage of Uganda's national parliamentary communication system.

First, we learned from the FFE that a nontrivial share of citizens, including especially marginalized citizens, want to communicate with their representatives in government using new technological innovations and are willing to pay to do so. This stands in contrast to accounts of disengagement reflecting alienation or apathy. We also know that many—although clearly not all (see app. fig. 8)—have the capacity and means to do so. The results from the FFE support the idea that mobile technology could, under the right conditions, change the relationship between voters and representatives in low-income countries. An examination of the scaled-up system alone would have masked this core insight.

Second, from an experimental manipulation in the NFE, we found little evidence that the differences across field experiments are due simply to scale effects. Specifically, we do not find evidence that system usage is a strategic response to how many others are contacting their MP via the ICT platform. We believe that improving our understanding of the conditions under which constituents might view IT-based communication with public officials as complements or substitutes is an important avenue for future research.

Third, complier analysis suggests that although the first stage of treatment using radio ads was not large, given the scale of the experiment it was far from being weak enough to account for the low engagement. Indeed, the estimated complier effect in the NFE is about 4% of the effect on usage observed in the FFE. Fourth, from the mechanism experiment we learned that there is a reasonably strong responsiveness to personal invitations to engage politically when interest articulation is at stake, but we do not find evidence of

the kind of differential responsiveness that would be needed to account for differences in flattening effects across experimental settings.

These findings suggest that the disappointing results of the uSpeak program are not driven by weak demand. In contrast, survey evidence suggests weaknesses in the marketing of the system itself. While a relatively large number of constituents were exposed to the radio ads, citizens had difficulty retaining and internalizing the information needed for acting conditional on hearing about the service. Moreover, our analysis suggests that agent effects (i.e., that the change in the identity of the implementer, which was easily observed by experimental subjects) likely have been very consequential. Specifically, we find strong evidence that general trust in the responsiveness of politicians is preventing engagement but is also rational. Interestingly in our case, agent effects do not stem from motivation differences between implementers (e.g., as identified by Berge et al. [2012]) but rather from the way agent identities interact with citizen expectations. In Uganda, as in many electoral authoritarian regimes—the most common regime type in Africa—low levels of political efficacy are discouraging political action; ICT innovations, by themselves, cannot force nonresponsive politicians to become responsive.

With the multiple pieces of evidence available to us, we infer that the failure of the nationwide program is not simply a function of weak demand on the part of citizens or to the weakness of marketing mechanisms but is a function of larger inequalities. Some of these, such as unevenness in receipt of invitations from Parliament, might be addressable through improved interventions. However, some reflect more fundamental weaknesses in the broader political system, most notably cynicism regarding the competence and motivations of politicians, which Parliament likely cannot address easily through technological innovation.

Our study also has broader implications for research methodology, especially for the extent to which outcomes of scaled-up programs can be gleaned from results of controlled small-scale interventions. The literature on scaling up has largely focused on assessing the extent to which experimental estimates in one context apply in another. Some of this literature highlights the problems in using a small handful of studies as the basis for inferences to different contexts (Open Science Collaboration 2015). Other work highlights the costs of extrapolation. Comparing nonexperimental and experimental estimates that rely on the same data, Pritchett and Sandefur (2013) conclude that nonexperimental estimates with the same subject population can better predict treatment effects as compared to experimental results from other contexts, because contextual variation can drive bigger differences

in the estimated effectiveness of a program than can selection bias.²⁴

Importantly, there should not have been great differences in the subject population between the FFE and the NFE. The FFE was offered to a random sample of subjects from every constituency in Uganda, while the scaled-up program was offered to a random subsample of 186 constituencies, out of a total of 238 (see app. fig. 4).

Thus, the differences we observe draw attention to a distinct problem largely overlooked by the extant external validity literature: the external validity across the nuts and bolts of interventions and not necessarily across populations.²⁵ This kind of validity problem is especially critical when lessons from carefully controlled small-scale studies are intended to inform policies to be implemented at a larger scale. Our results provide a cautionary tale for researchers and policy makers seeking to make such claims.

In our analysis, we identified several distinct reasons why outcomes of experiments may fail to replicate when brought to scale. These include already well-appreciated effects that relate directly to scale (see also Deaton [2010] on general equilibrium effects). In addition, we highlight possible effects related to the changing agents involved when interventions are implemented at scale (see also Bold et al. [2018] on capacity and motivation of implementing organizations), and we identify differences related to details in the design between controlled interventions and interventions implemented in the political wild, of the form that may be relevant for other studies.

Ironically when design details matter, a first response is to resort to controlled conditions to get those details right. This might be an appropriate approach when seeking to control for all factors but a manipulated variable of interest, but one core lesson from our study is that the importance of those details may only become apparent once researchers' control is removed.

24. There is a growing literature debating the trade-offs associated with different approaches to generating out-of-sample predictions based on experimental data. Hotz, Imbens, and Mortimer (2005) suggest using subject's observed characteristics as predictive of treatment effects independent of context. Gechter (2016) proposes a method that uses differences in outcome distributions for individuals with the same characteristics and treatment status in the original study and the context of interest to learn about unobserved differences across contexts.

25. Some of our divergent findings do relate to endogenous changes in populations as a consequence of the factors outside the control of the research team. For example, one reason we do not find price effects in the scaled-up program similar to those found in the controlled experiment can be attributed to the fact that the national intervention was taken up by relatively well-off and engaged citizens who are unlikely to be sensitive to a small price subsidy.

ACKNOWLEDGMENTS

We would like to thank the National Democratic Institute, especially Simon Osborn, Ivan Tibemanya, and Linda Stern, for a fruitful collaboration. We also thank Melina Platas and Jonathan Rodden for their generous support in facilitating our "mechanism" experiment in Arua and Jan Pierskalla and Laura Paler for allowing us to add questions about uSpeak to their nationally representative survey on oil in Uganda. This study benefited tremendously from comments from Michael Findley and Daniel Nielson and participants at Evidence in Governance and Politics's (EGAP) workshop meeting at Rice University and seminars at Johns Hopkins, Georgetown, and Harvard University.

REFERENCES

- Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of the Nazis in Prewar Germany." *Quarterly Journal of Economics* 130 (4): 1885–939.
- Allcott, Hunt. 2015. "Site Selection Bias in Program Evaluation." *Quarterly Journal of Economics* 130 (3): 1117–65.
- Arias, Eric, Pablo Balán, A. Larreguy Horacio, John Marshall, and Pablo Querubin. 2019. "How Social Networks Help Voters Coordinate around Information Provision to Improve Electoral Accountability: Evidence from Mexico." *American Political Science Review* 113 (2): 475–98.
- Banerjee, Abhijit V., and Esther Duflo. 2009. "The Experimental Approach to Development Economics." *Annual Review of Economics* 1:151–78.
- Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Kartika Sari Juniwyat, and Bertil Tungodden. 2012. "Business Training in Tanzania: From Research-Driven Experiment to Local Implementation." *Journal of African Economies* 21 (5): 808–27.
- Blair, Graeme, Rebecca Littman, and Elizabeth Levy Paluck. 2019. "Motivating the Adoption of New Community-Minded Behaviors: An Empirical Test in Nigeria." *Science Advances* 5 (3): 1–8.
- Bleck, Jaimie, and Nicolas van de Walle. 2013. "Valence Issues in African Elections: Navigating Uncertainty and the Weight of the Past." *Comparative Political Studies* 46 (11): 1394–421.
- Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur. 2018. "Experimental Evidence on Scaling Up Education Reforms in Kenya." *Journal of Public Economics* 168 (December): 1–20.
- Buntaine, Mark T., Brigham Daniels, and Colleen Devlin. 2018. "Can Information Outreach Increase Participation in Community-Driven Development? A Field Experiment Near Bwindi National Park, Uganda." *World Development* 106 (June): 407–21.
- Callen, Michael, and James D. Long. 2014. "Institutional Corruption and Election Fraud: Evidence from a Field Experiment in Afghanistan." *American Economic Review* 105 (1): 354–81.
- Christensen, Darin, and Simon Ejdemyr. 2020. "Do Elections Improve Constituency Responsiveness? Evidence from US Cities." *Political Science Research and Methods* 8 (3): 459–76.
- Coleman, Kenneth M., and Charles L. Davis. 1976. "The Structural Context of Politics and Dimensions of Regime Performance: Their Importance for the Comparative Study of Political Efficacy." *Comparative Political Studies* 9 (2): 189–206.
- Craig, Stephen C., Richard G. Niemi, and Glenn E. Silver. 1990. "Political Efficacy and Trust: A Report on the NES Pilot Study Items." *Political Behavior* 12 (3): 289–314.
- Croke, Kevin, Guy Grossman, Horacio A. Larreguy, and John Marshall. 2016. "Deliberate Disengagement: How Education Decreases Political

- Participation in Electoral Authoritarian Regimes." *American Political Science Review* 110 (3): 579–600.
- Dale, Allison, and Aaron Strauss. 2009. "Don't Forget to Vote: Text Message Reminders as a Mobilization Tool." *American Journal of Political Science* 53 (4): 787–804.
- Deaton, Angus. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48 (2): 424–55.
- Duflo, Esther, Rema Hanna, and Stephen P. Rya. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102 (4): 1241–78.
- Ferrali, Romain, Guy Grossman, Melina Platas, and Jonathan Rodden. 2019. "It Takes a Village: Peer Effects and Externalities in Technology Adoption." Unpublished manuscript. <https://bit.ly/2FvncCc>.
- Gechter, Michael. 2016. "Generalizing the Results from Social Experiments: Theory and Evidence from Mexico and India." Unpublished manuscript. <https://bit.ly/2JzWb15>.
- Green, Donald P., and Alan S. Gerber. 2008. *Get Out the Vote: How to Increase Voter Turnout*. 2nd ed. Washington, DC: Brookings.
- Grossman, Guy, Macartan Humphreys, and Gabriella Sacramone-Lutz. 2014. "'I Wld Like U Wmp to Extend Electricity 2 Our Village': On Information Technology and Interest Articulation." *American Political Science Review* 108 (3): 688–705.
- Grossman, Guy, and Kristin Michelitch. 2018. "Information Dissemination, Competitive Pressure, and Politician Performance between Elections." *American Political Science Review* 112:280–301.
- Grossman, Guy, Kristin Michelitch, and Marta Santamaria-Monturiol. 2017. "Texting Complaints to Politicians: Name Personalization and Politicians' Encouragement in Citizen Mobilization." *Comparative Political Studies* 50 (10): 1325–57.
- Grossman, Guy, Melina Platas, and Jonathan Rodden. 2018. "Crowdsourcing Accountability: ICT for Service Delivery." *World Development* 112 (December): 74–87.
- Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature* 42 (4): 1009–55.
- Heckman, James J. 1991. "Randomization and Social Policy Evaluation." NBER Working paper, National Bureau of Economic Research, Cambridge, MA. <http://www.nber.org/papers/t0107>.
- Hotz, V. Joseph, Guido W. Imbens, and Julie H. Mortimer. 2005. "Predicting the Efficacy of Future Training Programs Using Past Experiences at Other Locations." *Journal of Econometrics* 125 (1): 241–70.
- Kasara, Kimuli, and Pavithra Suryanarayan. 2015. "When Do the Rich Vote Less than the Poor and Why? Explaining Turnout Inequality across the World." *American Journal of Political Science* 59 (3): 613–27.
- LeVeck, Brad L., D. Alex Hughes, James H. Fowler, Emilie Hafner-Burton, and David G. Victor. 2014. "The Role of Self-Interest in Elite Bargaining." *Proceedings of the National Academy of Sciences* 111 (52): 18536–41.
- Malhotra, Neil, Melissa R. Michelson, Todd Rogers, and Ali Adam Valenzuela. 2011. "Text Messages as Mobilization Tools: The Conditional Effect of Habitual Voting and Election Salience." *American Politics Research* 39 (4): 664–81.
- Manski, Charles F. 2013. *Public Policy in an Uncertain World: Analysis and Decisions*. Cambridge, MA: Harvard University Press.
- Niemi, Richard G., Stephen C. Craig, and Franco Mattei. 1991. "Measuring Internal Political Efficacy in the 1988 National Election Study." *American Political Science Review* 85 (4): 1407–13.
- Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349 (6251). <https://doi.org/10.1126/science.aac4716>.
- Peixoto, Tiago, and Micah L. Sifry, eds. 2017. *Civic Tech in the Global South: Assessing Technology for the Public Good*. Washington, DC: World Bank.
- Pritchett, Lant, and Justin Sandefur. 2013. "Context Matters for Size: Why External Validity Claims and Development Practice Do Not Mix." *Journal of Globalization and Development* 4 (2): 161–97.
- Sheffer, Lior, Peter John Loewen, Stuart Soroka, Stefaan Walgrave, and Tamir Sheafer. 2018. "Nonrepresentative Representatives: An Experimental Study of the Decision Making of Elected Politicians." *American Political Science Review* 112 (2): 302–21.
- Sjoberg, Fredrik M., Jonathan Mellon, and Tiago Peixoto. 2017. "The Effect of Bureaucratic Responsiveness on Citizen Participation." *Public Administration Review* 77 (3): 340–51.
- Stokes, Susan C., Thad Dunning, Marcelo Nazareno, and Valeria Brusco. 2013. *Brokers, Voters, and Clientelism: The Puzzle of Distributive Politics*. Cambridge: Cambridge University Press.
- van der Windt, Peter, and Macartan Humphreys. 2016. "Crowdseeding in Eastern Congo: Using Cell Phones to Collect Conflict Events Data in Real Time." *Journal of Conflict Resolution* 60 (4): 748–81.
- World Bank. 2016. *World Development Report: Digital Divides*. Washington, DC: World Bank.
- Yanagizawa-Drott, David. 2014. "Propaganda and Conflict: Evidence from the Rwandan Genocide." *Quarterly Journal of Economics* 129 (4): 1947–94.