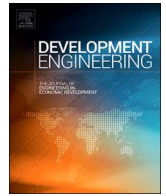




Contents lists available at ScienceDirect

## Development Engineering

journal homepage: [www.elsevier.com/locate/deveng](http://www.elsevier.com/locate/deveng)

# The double-edged sword of mobilizing citizens via mobile phone in developing countries

Aaron Erlich<sup>a</sup>, Danielle F. Jung<sup>b</sup>, James D. Long<sup>c,\*</sup>, Craig McIntosh<sup>d</sup>

<sup>a</sup> McGill University, Department of Political Science, Room 414 Leacock Building, 855 Sherbrooke Street West, Montreal, Quebec H3A 2T7, Canada

<sup>b</sup> Emory University, Department of Political Science, 1555 Dickey Drive, 327 Tarbuton Hall, Atlanta, GA 30322, USA

<sup>c</sup> University of Washington, Department of Political Science, 101 Gowen Hall, Box 353530, Seattle, WA 98195-3530, USA

<sup>d</sup> University of California, San Diego, School of Global Policy and Strategy, 9500 Gilman Drive, #0519, La Jolla, CA 92093-0519, USA

## ARTICLE INFO

### Abbreviations:

ICT Information and Communications Technology and Digital Media  
SMS Short Message Service  
USSD Unstructured Supplementary Service Data

### Keywords:

ICT4D  
Measurement  
Data Quality  
Political Participation  
Digital Inclusion  
Data Collection Methods  
South Africa  
Democracy and Governance  
Citizen Monitoring

## ABSTRACT

New innovations in mobile technology provide an unparalleled opportunity for researchers and organizations to scale communications with citizens in the developing world, but bring new challenges in terms of how to generate and retain engaged users. We report on a number of technical dimensions based on our experience building a bi-directional multi-channel mobile phone platform to engage citizens in South Africa's 2014 presidential election. Specifically, we deployed the "VIP:Voice" platform at national scale to conduct opinion polling, to allow citizens to report on political activity, and to engage citizen monitors for polling stations on election day. Our platform operated across multiple device types, from flip-phones to Twitter, and consequently provides critical lessons on the most effective means of gathering and disseminating a rich variety of data depending on the user's device type. We compare different means of obtaining location in the absence of GPS, and show how different formats for soliciting and entering data generated very differential response rates. Our paper illustrates a number of concrete ways in which platform development driven by smartphone logic does not translate easily for users of more basic mobile phones, including whether questions are presented passively in a menu or pushed to a user's phone, and the format in which user data are entered. This paper is intended to provide actionable guidance for researchers and organizations deploying ICT platforms to interact with citizen users at a national or cross-national scale in international development.

## 1. Introduction

The adoption and expansion of information and communications technology and digital media (ICT) has radically changed the scale at which researchers, activists, and organizations harness technology across sectors in international development (e.g., Aker et al., 2012; Blumenstock et al., 2015b). However, stakeholders and platform developers confront important trade-offs in using mobile and digital tools to improve program design, implementation, and evaluation. One critical issue confronting researchers is scaling development interventions to reach as wide a population as possible (Heeks, 2008; Sachs and McArthur, 2005; Tomlinson et al., 2013), while maintaining a stable user base that provides high quality data. A second area of focus involves weighing deploying ICT platforms on a single mobile or digital channel and therefore tightly controlling the medium of communication while limiting use, compared to multiple channels, which expands the potential sample of users but decreases control over the system.

In this paper, we address these issues by reporting on several

central features of "VIP:Voice," a multi-channel ICT platform that we designed and deployed during South Africa's 2014 national election. We highlight how the platform's cross-channel engineering performed in providing robust data from a diverse population of citizen users. We built VIP:Voice to reach, recruit, and engage South Africans in several modes of political participation, including sharing their opinions about political topics, reporting election-related events, and monitoring voting districts (colloquially known as polling stations, the term we use here) on election day. Users could interact with the platform through five ICT channels: USSD/SMS, Mxit (a widely used South African Facebook competitor), Mobi (mobile web), Google Talk (GTalk), and Twitter. The platform included embedded randomization protocols to study experimentally the effects of incentives and message content on uptake, registration, and usage.

We first document the advantages of using such a system for low-cost scalability and broad representativity, as well as fine-grained analysis across and within users' message-level interactions. Second, we discuss the challenges the platform confronted regarding programming

\* Corresponding author.

E-mail address: [jdlong@u.washington.edu](mailto:jdlong@u.washington.edu) (J.D. Long).

randomization into messaging content, eliciting geospatial information from users, and retaining participation. Data were sent over multiple channels compatible with a broad set of devices, including standard and smart phones, tablets, and computers. This diversity of technological channels allows us to provide some important lessons learned in terms of implementing complex spatial and non-spatial randomized messaging treatments using an open-source bi-directional technology application for sending and receiving information from end-users.

The paper's findings demonstrate that multi-channel platforms provide researchers an exciting opportunity to scale their research. VIP:Voice's cross-channel design allowed it to reach a much larger and broader cross-section of the South African population than it would have using any one channel. On the downside, we find eliciting geospatial information proved extremely difficult, standardizing messaging across platforms with different character limits required significant development time and effort, and that ICT platforms present some logistical challenges when we attempt to build complex randomization protocols into the dispatching of messages from the system to users. As reported in Ferree et al. (2017), we discuss the high rate of attrition from the platform, as well as implications of this reality for such platforms moving forward. We encourage future work to understand how to make cross-channel interventions and experiences more comparable. Many of the opportunities and challenges we faced with VIP:Voice and report here are important for not only development engineering, but also robust program implementation and evaluation.

Our study makes several contributions to the design and evaluation of ICT platforms in international development. First, we provide insights to those studying the dynamics of “digital inclusion” in developing contexts (Madon et al., 2009; Walsham and Sahay, 2006; Warschauer, 2003), and how initial engineering decisions define the population from which potential users are sampled. While developing countries have experienced an explosion of ICT innovations and user bases over the last decade, previous studies note that across technologies, income, and socio-demographic features, certain types of users may be systematically over or under-represented in these platforms (DiMaggio et al., 2001; Thompson, 2008). Subsequently, these diverse populations may interact with technology in distinct ways, producing important differences in digital participation among those who are “included.” While many previous projects target ICT users identified by program pre-registration, such as health-workers or farmers, we built our platform for no pre-defined user base; this design allowed us to recruit *any* South African with a phone to join the platform. Participation over cheaper and easy to access channels (i.e., SMS/ USSD) generated users representative of South Africa's more excluded populations, but at the cost of certain technical difficulties and design limitations. Conversely, participation over digital channels on social media brought in users more typically included in digital platforms at the benefit of fewer technical difficulties and design limitations. Thus, our results highlight an important tension between technical design considerations and population/user recruitment; we use a cost-effectiveness analysis to demonstrate clearly how marginal recruitment costs and representativity play off against each other across channels in our context. Our overall platform is substantially more representative of the nation than it would have been if it had been implemented over any single channel.

Second, our results provide insights at the intersection of applied social science research methodology and computer science research on ICT platform design and development. Research in this area has addressed problems of citizen welfare across development sectors like agriculture (Aker, 2011), banking and mobile money (Shaikh and Karjalaoto, 2015), education (McEwan, 2015), and health (Källander et al., 2013; Rajput et al., 2012). Our study more narrowly falls within recent technological innovations in corruption and political accountability (Bussell, 2010; Grossman et al., 2014; Humphreys and Weinstein, 2012), a part of larger concerns with citizen monitoring, reporting, and participation in governance and elections (Bailard,

2012; Bailard and Livingston, 2014; Paluck et al., 2010). While previous research using ICT platforms in these spaces has noted the problem of uptake and attrition over time (Findley et al., 2013; Grossman et al., 2014), we explicitly demonstrate the challenges of gaining and maintaining participation for research endeavors with little or no face-to-face contact with the study sample. Our context is potentially special, however, in that we are attempting to build and deploy a national platform quickly for a specific event (the election), and experiences with participation may differ when the interaction with users is not time-bound (e.g., Chicoine and Guzman, 2017; Dhaliwal and Hanna, 2017). These challenges include attrition as the result of asking for locational information as we show here, and, as shown in Ferree et al., (2017), attrition as a result of the system not consistently provide registrants material or financial benefits (although incentives were offered at different stages of the project to a sub-set of users). The governance sector therefore presents a “hard case” to evaluate the effectiveness of this approach: compared to areas where users may receive a direct and immediate benefit from engaging with an ICT platform on multiple reporting activities, governance interventions deliver fewer immediate private benefits.

Lastly, we contribute specific technical and research design insights to the creation, roll-out, and evaluation of development engineering platforms (Brunette et al., 2013; Hartung et al., 2010; Okolloh, 2009). Technologists have laid out minimum requirements for development engineering platforms; Hartung et al., (2010) argue “information services must be composed by non-programmers, deployed by resource-constrained organizations, used by minimally-trained users, and remain robust despite intermittent power and connectivity” (1). However, there may be other important requirements driven by the project's social and research goals. Specifically, we show the development and engineering process must also pay attention to population representativeness of study samples, geographic information collection, and messaging/text comparability. While these issues are no doubt salient to social scientists, they are not automatically at the forefront of decisions over platform design by applied computer scientists, who are, understandably, focused on technical features. Our experience underscores the necessity of integrating more fully from the outset the insights and contributions of computer scientists with social scientists and development practitioners. These partnerships can generate linkages that will contribute better research designs and program implementation to improve and scale platforms.

This paper proceeds as follows. In Materials and Methods (Section 2), we lay out the design (Section 2.1) and implementation (Section 2.2) of VIP:Voice around a citizen mobilization campaign during the 2014 South African elections. In Results (Section 3), we first discuss cross-channel comparability of responses and response rates (Section 3.1), and then look at the relative successes of different methods to elicit geospatial information from users of phones without GPS (Section 3.2). In Section 3.3 we describe several technical issues researchers interested in carrying out similar experiments would likely face. Finally, we describe the relative costs effectiveness of development and implementation across channels, with an eye to generating a representative sample across the platform as a whole (Section 3.4). We conclude by offering a set of recommendations for researchers and practitioners interested in deploying multi-channel ICT platforms.

## 2. Material and Methods

This section describes the design features of our ICT platform, VIP:Voice, and how it engaged South African citizens during the 2014 national election campaign.

### 2.1. Platform Design: VIP:Voice

We launched VIP:Voice to reach, recruit, and engage users around South Africa's 2014 general election, while simultaneously running

Channel	Description	Access via			
		Basic	Feature	Smart	Web
USSD	Mobile Technology to Send/Receive Texts	✓	✓	✓	
Mxit	South Africa's Largest Social Network		✓	✓	✓
Mobi	Mobile Web for Internet-Capable devices		✓	✓	✓
Gtalk	Online Instant Messaging Service		✓	✓	✓
Twitter	Online Social Networking Service		✓	✓	✓

Fig. 1. Channel description and accessibility requirements.

social science experiments and collecting observational data and meta-data on users. Different types of users interact with ICT on a myriad of devices. We therefore set out to launch a platform to appeal to the broadest possible cross-section of South Africans. We worked with Praekelt, a South African technology firm, to design our platform using their application Vumi, and its hosted version Vumi-Go, as our chosen solution. Vumi is an actively developed open-source application, referred to as a “messaging engine.”<sup>1</sup> Vumi operates over different mobile phone providers, but also targets a broad section of the population by utilizing a wide range of modes of communication, including not only USSD and SMS, but also social media and web applications.

Compared to applications such as Frontline SMS, Vumi's innovation includes the ability to send and receive messages from *multiple* ICT channels and store all interactions in one data store. It uses a dispatch technology to enable the pushing and receiving of messages across different telecom networks and types of communication protocols. Vumi's data store also contains rich meta-data, including information on the timing of message dispatch and receipt, along with the content of each message. The use of a local platform also means that users are responding to a local telephone number resulting in substantially lower user SMS costs than if the response must be sent to an international phone number.

Many studies build user bases for ICT platforms from extant databases of previously identified potential users, such as those who pre-register for a program advertised or deployed through the platform (like a health intervention with community health workers or election monitoring program with a network of civil society organization members), citizen users identified from household surveys, or other datasets of phone numbers from registrants provided in bulk through purchase from telecom or marketing companies. Instead, VIP:Voice obtained participants directly from the overall population using a unique contact and registration method (described below), rather than any previously gathered registration data, phone numbers, or organizational structure. While this method presented significant operational challenges that we address below, it also meant that every South African voter, on any device, with the ability to receive and read an SMS message could potentially enter the study sample. Accordingly, our study provides a robust proof of concept on purely digital recruitment into large-scale studies. Within the system, we employed a series of experimental and non-experimental methods to recruit and engage users, push and pull information, obtain survey responses, and deploy “feelings thermometers” to measure political attitudes in nearly continuous time in the weeks prior to the election.

<sup>1</sup> Available at <https://github.com/praeekelt/vumi>. Praekelt originally developed Vumi to address a variety of development problems in South Africa, including in public health and youth employment.

## 2.2. Project implementation

We rolled out VIP:Voice over four phases: (1) registration; and engagement (2) before, (3) during, and (4) after the election. We initiated enrollment in Phase 1 one month before the election. Users could interact with the platform through five ICT channels: USSD/SMS, Mxit (a widely used South African Facebook competitor), Mobi (mobile web), Google Talk (GTalk), and Twitter. Standard phones lacking internet functionality required interaction via short message services (SMS, or text messages) and unstructured supplementary service data (USSD), an interactive text-based system that can reach users of all types of phones. USSD is an SMS alternative for GSM-based phones. USSD has been explained as an “interactive SMS” (Wouters et al., 2009, p. 5), where multiple, pre-coded questions are sent to respondents in one session and users answer in real time on an open network session. Mxit was at the time South Africa's largest social network and works on feature and smartphones; Mobi is a provider of mobile web smartphone platforms; GTalk and Twitter could be accessed by feature or smartphones. Fig. 1 displays the capabilities by channel of development.

### 2.2.1. Recruitment

We employed a channel-specific set of recruitment activities and spent varying levels of resources on different channels. Given widespread penetration of mobile telephones in rural areas and informal settlements in South Africa, where other digital media does not penetrate, we heavily targeted SMS/USSD interactions. We also spent heavily on Mxit, which achieved the highest uptake. Splash ads and banners advertised our platform on social media channels (Twitter, Mxit, and Mobi (see Appendix)). We recruited people to our SMS/USSD channels primarily through advertising with Please Call Me (PCM) messages. Provided by telecom companies, South Africans send an average of 14 million overall unique PCMs per day. A sender texts a PCM to a recipient, requesting a return phone call. The recipient of a PCM sees the number requesting a call as well as an ad. Advertisers pay for PCMs, not senders or recipients.<sup>2</sup>

The total recruitment effort resulted in about 263,000 individuals contacting the platform, 134,047 responding to the initial engagement question, and 90,646 completing the Terms and Conditions. Just under half of registrants entered through the PCM-linked USSD channels; a similar number entered via Mxit. The USSD and the Mxit channels contained almost 94% of registered users. The remaining participants entered through Mobi, print advertising, Gtalk, or Twitter (see Ferree et al. (2017) for additional details).

### 2.2.2. Participation

In Phase 2, the platform invited registered individuals to provide their demographic data and report on election-related events. Users could report or enter data at any time, but reports were solicited also with information “pushes” (that provided summary statistics on data collected to date from other VIP:Voice participants) and “pulls” (randomized incentivized nudges and reminders to participate) leading up to election day. Prompts to engage were done to gather data on various types of activities via the platform dashboard (see Appendix) and through one-directional SMSs sent on a regular basis. A goal of Phase 3 was to evaluate whether ICT could enlist citizens to observe and report on election day events and outcomes at polling places. Phase 4 implemented a Get Out the Vote (GOTV) experiment and additional surveys to gauge satisfaction with the electoral process (incentivized with a lottery).

As shown in Ferree et al. (2017), the different channels on the

<sup>2</sup> We purchased ad space for VIP:Voice for 49.8 million PCMs to recruit subjects. We randomized the PCM message with a “standard” arm encouraging registration, but users pay full messaging costs; a “free” arm with no interaction fees; and a “lottery” arm offering a chance to win R55 (approximately 5.25 USD at the time).

platform yielded very different user groups. Those coming in as USSD users, who primarily used more basic flip phones, are 65% female and 94% Black. Mxit users, by contrast, are 62% male and only 82% Black, with 14% of users who provided demographic information self-identifying as Coloured, or mixed-race. Users of the USSD channel were almost 4 years older on average than Mxit users. This demonstrates that a multi-channel approach can be a critical component of yielding a more representative user base than would be generated by a smartphone-only platform.

### 3. Results

#### 3.1. Comparisons across channels

A central component of our project explicitly required multiple ICT channels that could reach and incorporate various types of users across South Africa. This approach promotes digital inclusion to expand the potential user base of an ICT intervention. In this section, we discuss three advantages of using multiple ICT channels unified into a single platform.

##### 3.1.1. USSD versus smartphone channels

There are two important reasons to use USSD alongside SMS. First, USSD can lead to large cost decreases over SMS because the per-session costs are generally much lower than the cost of SMS. Second, USSD yields much cleaner and more usable data as respondents can only give pre-coded answers that correspond to numbers on their phone keypad. Conversely, SMS data are often hard to read, interpret, or systematically categorize, especially when dealing with many users who may each send unique responses because users respond with free-form text. As shown in Table 1, while SMS data can be cleaned to achieve usability rates comparable to USSD in large-scale studies, this requires some combination of machine learning techniques, regular expressions, or hand coding. To achieve the response rates shown in Table 1, we hand coded all the 189 unique responses to a question we sent out via SMS which was similar to a question respondents answered over USSD and achieved a 4.46 percentage point lower rate of usable data with SMS compared to USSD. However, regardless of whether data are usable, USSD data will, by default, be much cleaner and create fewer problems in subsequent analysis.

There are several important considerations when using USSD. First, because smartphone app developers typically visualize menu structure with a web-based model in mind, they may overlook how tedious and slow it is to access deeper branches in a decision tree on a small-screen flip phone using the timed interaction of USSD. Instead, users of more basic phones may be more likely to engage with content if it is directly pushed to them in the form of an SMS message rather than requiring them to navigate through successive layers in a decision tree to access it. Therefore, a central question for researchers is how channel medium affects response rates in a menu driven system versus a system in

**Table 1**  
Examples of freeform SMS and structured USSD questions.

<b>Free-Form SMS Question:</b>	
<i>VIP: Voice wants 2 know if election held 2moro what party would u vote for: 1. ANC 2. COPE 3. DA 4. EFF 5. Other 6.not vote 7. Skip SMS cost std rates</i>	
Number of responses:	4,631
Number usable responses:	4,277
Percent responses usable:	92.36%
Number of unique answers:	189
<b>Structured USSD Question:</b>	
<i>Which political party do you feel close to? 1. ANC 2. Agang 3. COPE 4. DA 5. EFF 6. IFP 7. Other 8. I don't feel close to a party 9. Skip</i>	
Number of responses:	10,531
Number usable responses:	10,196
Percent responses usable:	96.82%
Number of unique answers:	10

**Table 2**  
Differential Response Rates to "Pushed" versus "Menu" Question Formats.

	Response Rate to "Pushed" Questions (1)	Response Rate to "Menu" Questions (2)
Differential Rate for USSD user	0.111*** (0.004)	-0.116*** (0.003)
Response Rate among Mxit users	0.0229*** (0.001)	0.140*** (0.002)
Number of observations	47,682	47,682
R-Squared	0.04	0.02

OLS regressions with robust Standard Errors, regression estimated within the population of registered users of USSD (predominantly non smartphone) or of smartphones (Mxit). \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

which questions are sent to users. Second, USSD development requires more up-front work than SMS development because of its interactive nature, offsetting some of the costs gains attained by using the technology. A third potential concern with USSD involves variation in the length of time the connection to the network remains open during the USSD response session.<sup>3</sup> These rates vary by country, and potentially by provider, requiring care to understand the timeout environment.

Table 2 demonstrates that users on USSD have much higher response rates when pushed a question over SMS as compared to Mxit users who were pushed a question natively within the Mxit environment. On the other hand, Mxit users are much more likely to respond to menu-based question compared to those on USSD.<sup>4</sup> This finding suggests that in multichannel platforms, those on USSD channels will need to receive questions as direct requests for information sent via SMS rather than having questions sit within passive decision tree structures, relative to those using more content-rich phones and interacting over social media channels. Sending reminder messages over SMS will offset some of the cost savings accrued by developing USSD applications (compared to SMS-only).

Compared to many prior studies, our recruitment method follows a radically different approach. Many ICT projects distribute or "seed" their own devices to a defined set of users. This approach standardizes the device on which users send and receive messages, giving the implementer maximum control over the application and use in small, tightly controlled interventions and in traditional sample survey enumeration (e.g., Callen et al., 2016). For example, a study of health workers could deploy hand-held devices to facilities or staff (Pop-Eleches et al., 2011) similarly, projects with citizen participation could lend devices to user groups like farmers, veterinarians, or market stall vendors (Rezaee et al., 2015).

Our approach highlights two interrelated reasons to depart from such a design: marginal cost and scalability. Allowing end users to engage with their own (previously procured) devices drastically reduces the marginal cost of communication. In terms of costs, our platform does not require implementers to pay for the users' technology and, all training, theoretically, can be carried out over the platform itself. Similar to platforms like Amazon's Mechanical Turk (Benoit et al., 2016) for data coding, such ICT platforms that embrace technologies

<sup>3</sup> In South Africa, cell phone operators limit the overall amount of time for USSD response to 45–60 seconds — or about four questions — before the operator closes the session (the session times out). If this occurs, a user must dial a USSD short code to re-enter the system.

<sup>4</sup> While any user of Mxit must be using a feature phone or smart phone that is web enabled and features a larger, easier-to-use screen, some of the users who entered the platform as USSD users also may be using smartphones. Hence the actual distinction between response rates would likely be even larger if we were able to cleanly segregate users by phone type.

such as USSD or social media channels do not change the functional form of the communication costs (they are still linear) but significantly reduce the marginal cost as to make the scale of large magnitude projects feasible. Finally, this approach to ICT allows scalability because the human capital necessary to seed devices, and for users to achieve familiarity with new device, also drastically reduces costs and time.

When scalability at low costs is a consideration, projects like VIP:Voice sacrifice a degree of implementer-side control gained by careful distribution and management of devices. While such projects pay upfront development costs for a platform that can engage with users across a wide variety of channels there are major downstream advantages of such systems where users can participate on their own devices. In other words, this type of system is high in up-front development costs, but adding users does not require additional device purchase or training. It is also likely to decrease users' cognitive costs related to using a new device or application— even more so because users interact primarily through a channel of their choice, whether social media or USSD.

An important consideration regarding reaching people on their personal phones involves finding the best way to communicate with them. A large cost component of any ICT platform is SMS, the most commonly used method of communication for informational campaigns. Recent feasibility studies and randomized control trials examine the use of SMS technology to carry out both bi- and uni-directional studies for a wide variety of public health problems including alcoholism and alcohol consumption (Kuntsche and Labhart, 2013; Kuntsche and Robert, 2009), reproductive health (Merrill et al., 2013), and sexually transmitted infections (STIs) (Hardy et al., 2011; Lim et al., 2008; Swendeman and Rotheram-Borus, 2010). For instance, the use of SMS has become commonplace in medical studies in order to remind patients to attend appointments, prompt them to take medication, or to provide short educational messages (Lim et al., 2008; Pop-Eleches et al., 2011).<sup>5</sup>

For deploying a large-scale platform such as ours, SMS has several downsides, including cost, recruitment, and the number of questions it is possible to ask. First, some individuals in developing countries could incur costs that would prohibit them from sending regular or consistent SMS messages.<sup>6</sup> While prices have fallen dramatically in recent years, pre-paid airtime means every text requires one cost unit more in pricing. While there are increasingly discounts for subscription or bundled SMS plans, only wealthier people have addresses and automatic billing options for subscription as opposed to pre-paid airtime or enough liquid capital to purchase bundles. In many studies with pre-defined users recruited specifically for the study or registered via some other activity (like hospital registration), the SMSs often involve just one uni-directional exchange (such as a reminder from the hospital), and are often cost-free, as the medical service provider pays the costs of SMSs. In distinction to such a usage pattern, we desired a high-level of baseline involvement in our study and wanted users to avoid high marginal SMS costs.

A second problem using SMS is that messages are “one-off,” which make them a poor vehicle for asking multiple survey questions and any follow-ups, particularly in a short time frame and answering multiple questions can be cumbersome. Phones often display these messages in chains and individuals may not see the multiple messages they received. SMS data is returned as text, often containing many typos and errors, although new machine learning techniques can aid in the analysis of such data (see Roberts et al., 2014).

<sup>5</sup> While in principle, users could send SMSs back to say they were going to attend their appointments, many studies use uni-directional contact.

<sup>6</sup> Although noted in Ferree et al. (2017), many users still engaged in VIP:Voice despite having to pay usage costs.

### 3.1.2. Cross-channel compatibility in response rates

Social scientists understand the importance of careful attention to survey question wording. Changes in the wording of questions can dramatically alter the ways respondents understand them, and therefore, the distribution of the frequency of answers (Schuman, 1996). Changing wording can impact not only opinion questions but also factual ones (Kalton and Schuman, 1982). Because question wording matters, many researchers insist on standardizing question wording across channels, as channel-specific versions of the same question may introduce unquantifiable differences in measurement of the underlying phenomena across channels. Collecting survey data through Vumi-Go raised the difficulty of asking survey questions more typically administered in internet or face-to-face surveys under the extreme character limits imposed by SMS and social media.

While social research has historically worried about how mode of survey administration affects survey response (face-to-face, paper-based surveys, telephone, mail, or internet surveys) (e.g., Dillman et al., 2009), none of these survey modes demand extreme character limits, as required by SMS, USSD or Twitter. Character length impositions prohibit longer instructions, context, or help often present in other types of self-administered surveys via web or mail. Therefore, the designers of a multi-channel platform need to decide whether to standardize all questions to the number of characters allowed by the most restrictive technology, or allow text length to fluctuate depending on channel-specific character restrictions.

To complicate matters even further, South Africa, like many developing contexts, is a multilingual environment. We deployed parts of our platform in the three most spoken of South Africa's 11 official languages. Therefore, our platform's messages needed to conform to the character count not only in English, but also in Afrikaans and Zulu, all of which use the Latin alphabet. In our study, this caused problems because some languages require more characters than others to convey the same message. A challenge we did not encounter—but many other researchers will likely face—will involve cases where a mixture of Latin and non-Latin character sets are used. Indeed, most non-Latin writing systems for SMS (e.g., Arabic, Japanese) only allow 70 characters because SMSs are based on bits (binary code), and Unicode require more bits to encode each character.

We opted to standardize our platform's messages to make all communication the same. As a result, to meet character limits, we used many conventional abbreviations and shorthand developed from text and chat. However, these abbreviations and shorthand are not equally developed across languages, making translation protocols important for such endeavors. Additionally, abbreviations and shorthand may not be equally well understood across age groups and genders.

Strict character limits raise notable challenges in providing instructions and information for subjects in our study. In Phase 3, we had to explain to our users recruited to serve as election monitors that we required them to go to their polling station both on election day and the day after. However, as practitioners who worked with Vumi-Go explained to us, users in ICT platforms generally will look at very few instruction screens or messages. Therefore, we had to downsize considerably instructional text to meet character limits and not show multiple screens. Table 3 shows one example of this downsizing. Such downsizing of text can strongly affect subject's understanding of the task.<sup>7</sup>

To understand how string length affects response, we use the variation in the number of characters used in the questions that we do have (even with our text limits) to see if the length of the text is correlated with response rate. We expected longer questions to elicit lower response rates both because they were likely to take more time to

<sup>7</sup> It may also be that the understanding of the task will be correlated with age, since younger people tend to be more versed in text message abbreviations.

**Table 3**  
Original, Revised, and Final version of Phase 3 question.

First Draft (Non-Shortened)	Second Draft (Shortened)	Final (Meeting Character Limits)
Thanks for agreeing to be a citizen observer! You will be asked to report on ur experience on election day and to photograph the tally on the day after election day. All activities are within your rights as a SA Citizen! Make your reports to tell the world, and learn about what others report! Pls be sure to think of your safety first and leave any tense situation.	Tnks! Ull b asked 2 report on eday and PHOTO the results the day after eday. Evrythng u do is within ur right as an RSA citizen. Make ur reports 2 tell the world & learn wut others report! Pls b sure 2 think of ur safety 1st!	Read T & Cs & accept 2 join. Ull b asked 2 report on 7 & 8May. Evrythng u do is within ur right as a citizen. Regents of Uni California(UC) doing study on elections in SA.

**Table 4**  
How does question length affect response rates?

	USSD only (1)	Mxit only (2)	USSD + Mxit, Interaction Effect (1)	Question & Language-Specific FE (2)
Question Length	-0.0000081 (0.000006)	0.0000192 (0.000053)	0.0000192 (0.000053)	-0.0002460 (0.000336)
Length* USSD			-0.0000272 (0.000053)	0.0000261 (0.000044)
USSD Sample			-0.176*** (0.006700)	-0.185*** (0.005350)
WhatsUp Question	-0.0104*** (0.000308)	-0.0864*** (0.003340)	-0.0864*** (0.003320)	
WhatsUp* USSD			0.0760*** (0.003330)	0.0748*** (0.003160)
Constant	0.0234*** (0.000638)	0.199*** (0.006710)	0.199*** (0.006670)	0.199*** (0.032000)
Observations	703,152	627,984	1,331,136	1,331,136
R-squared	0.001	0.009	0.082	0.082

Unit of analysis for the table is the question/individual. OLS regressions with Standard Errors clustered by question/language/channel, regression estimated within the population of registered users of USSD flip phones or of smartphones (Mxit) \* p<0.1, \*\* p<0.05, \*\*\*p<0.01.

read, but also because they were likely to include more abbreviations. We assign each user the *length* associated with the question in the language in which they received their communication. We analyze two different sections of our platform — the Answer & Win section, which contained key demographic questions and the VIP Question section, which pertained to the elections. As seen in Table 4 and Fig. 2, contrary to expectations we do not find any relationship between the Question Length and question response rates. The lack of relationship is robust to alternative model specifications in which we pool all of the different questions across the Answer & Win and VIP Question sections, and to the inclusion of question language fixed effects.

Our results suggest that, even on USSD where users are typically interacting using small screens that make the display of long messages cumbersome, response rates do not respond to question length. However, Mxit response rates were, on average higher for both WhatsUp and VIP questions. Still, we caution researchers to think carefully about the trade-offs between standardization and user experience. UX (User experience) developers focus on optimizing the experience for their users in a platform consistent manner. Such optimization, however, is often not in line with experimental or observational research goals of maintaining the exact same treatment across channels. This problem is particularly acute if there is an interactive effect between channel and treatment. In our study, we decided to standardize, but this choice likely may have led to some misunderstandings on the part of our respondents.

Additionally, we think our results highlight two areas of research critical to understanding better how meaning is conveyed through cross-channel research such as our own. First, scholars have carried

out few, if any, experiments on how internet and text-based abbreviations affect response rates. We believe theory should drive these experiments in order to clarify under what circumstances abbreviations alter respondents’ understanding of questions. Second, while our data show that channel matters, we do not know if individual responses vary on average if the respondent receives the same message via different channels (e.g., Twitter direct message or SMS).

### 3.2. Elicitation of geospatial information

Next, we highlight important challenges associated with a multi-channel platform like VIP:Voice. We discuss our lessons learned to deepen the knowledge base of scholars and implementers in the field of development engineering. To do so, we explain the challenges we faced in platform design and rollout, as well as the solutions we employed to address these issues. Our goal is to provide this evidence for other researchers in international development so they can save time, decrease costs, and build on our work to design and implement more elegant solutions.

The collection of geospatial data raises both practical and ethical dilemmas. Practically, while smartphone usage is on the rise, most South Africans still do not possess a smartphone (according to Pew, in 2016, 37% of South African adults have a smartphone, 52% have non-smart cell phones, and 10% have no mobile phone).<sup>8</sup> Only smartphones enable applications to track user location.<sup>9</sup> However, for a smartphone to track location, the user needs to turn on location services and agree to allow the app to track location. According to our South African partners, many wealthier South Africans turn off location services because of repeated stories and warnings in the media about sophisticated robbers using mobile phone GPS location data to commit carjackings and other types of theft (for an exposition of this hypothesis, see Chen et al., 2008, p. 10). Moreover, even if it were possible to track users via GPS location, academic research projects operate under stronger protocols of informed consent than do commercial actors; collecting mobile geolocation data de-identifies most subjects because their movement patterns are unique (Shilton, 2009).

Many ICT platform designers who want to locate users encounter daunting challenges when interacting with those who are unwilling or unable (or do not have smartphones) to share their location. Before our study, no projects implemented through Vumi-Go had attempted to carry out geospatial data collection of their users according to Praekelt. Absent users’ devices automatically sharing their location, platform designers must rely on users to input their own locations. However, obtaining high quality data from user input is difficult.

In our case, we needed geolocation for two reasons. First, for the entire population of registered platform users in Phase 2, we desired to

<sup>8</sup> <http://www.pewglobal.org/2016/02/22/smartphone-ownership-rates-skyrocket-in-many-emerging-economies-but-digital-divide-remains/>.

<sup>9</sup> A standard phone without GPS services can only be located by the last cell tower with which it interacted; such geolocation is much less accurate than GPS. Moreover, these non-anonymized cell tower data are generally unavailable to researchers, though, in a few exceptional cases, researchers have obtained anonymized data (Blumenstock et al., 2015a).

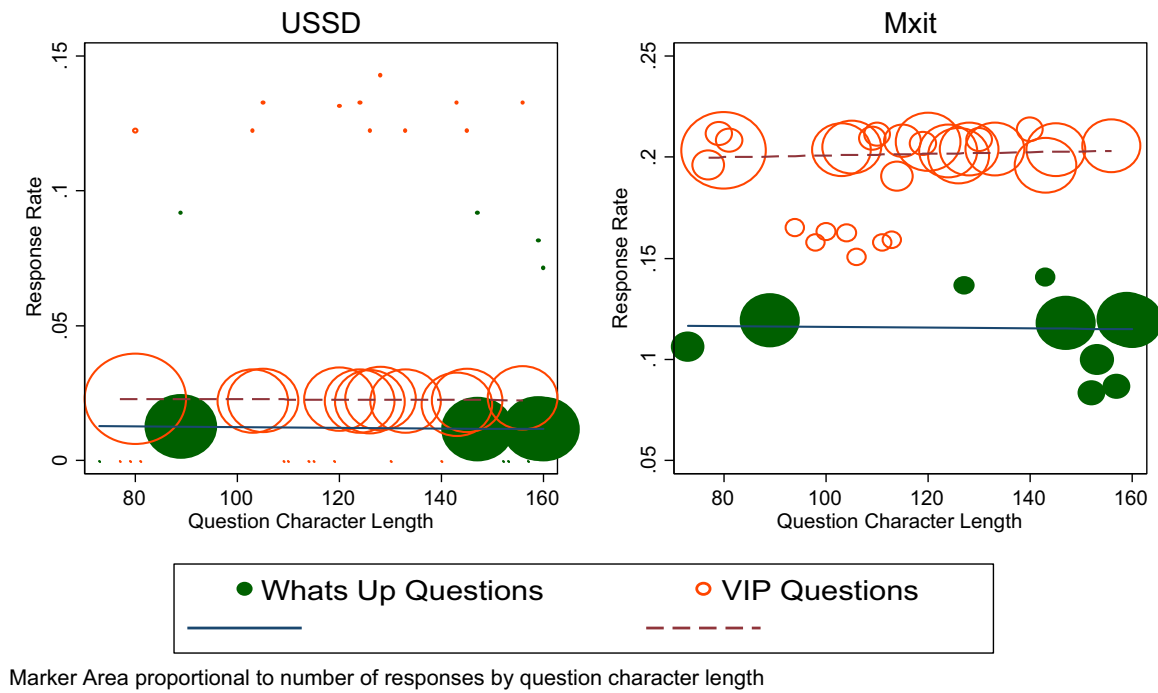


Fig. 2. Response rate by question length and channel.

conduct an experiment at the level of the polling station. Given our interest in understanding the effect of various messaging campaigns and linking these campaigns to behavioral voting outcomes, we wanted to randomize the messaging through our platform on the level of polling station, the lowest level of aggregation for which South African authorities collect data.<sup>10</sup> These research objectives required us to identify the polling station where each user was registered to vote. Second, in Phase 3, from within the population of users of our platform, we aimed to recruit individuals to participate in election monitoring, and randomly vary the monetary incentives of our recruited election monitors to estimate how incentivization shaped platform users’ willingness to participate. This deployment of platform users as election monitors also required assigning each user to the polling station in which they were registered.

We encountered two concrete obstacles in eliciting users to input their geographic location. First, users may reasonably refuse to share geographic information unless a clear justification is given to them as to why the platform seeks to acquire this information. Second, absent automatic GPS data from a smartphone (assuming they are responding from within the spatial boundaries of their voting district), user-entry of text is an inherently error-prone process. Furthermore, this error-prone, user-entered data then needs to be matched against a universal database of locations (whether it is a list of spatial units such as counties, or a residential address in countries in which mailing addresses exist).<sup>11</sup>

We describe our original methodology, designed for use in both Phase 2 and 3 to capture fine-grained geospatial information at the level of polling station based on user-entered input. We discuss the limitation of using the Google API, the back end we used to match user input

against a database of locations. Mainly, the limitations were not related to the Google API but rather were a function of low quality user input data; moreover, users attrited from the platform when asked for their geographic information. We show estimates of the percentage of usable address information we obtained before we shut down this user-input method and demonstrate the type of usage errors that individuals make in such a context. After our original method failed, we developed a new approach for Phase 3. We explain our solution, which involved eliciting sequentially more fine-grained data about individuals’ location.

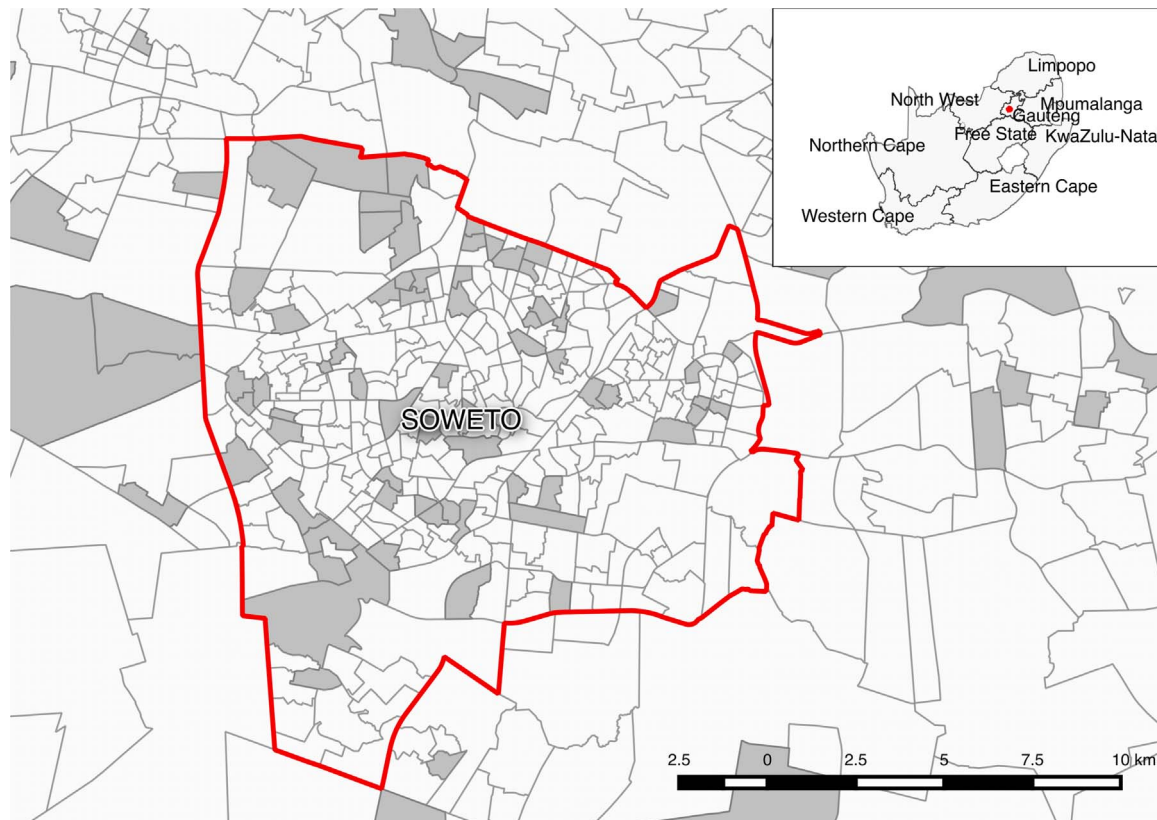
In the study’s original conception, we intended to obtain users’ home addresses to place them within the boundaries of their polling station precincts (not to rely on GPS location data). Understanding that it was not advisable to ask for location data without providing a reason, we formulated our request to highlight how we would use the data. After users registered and we prompted them with a teaser question to drive engagement, we asked: “Thanks 4 joining! 2 begin we need ur voting ward. Reply with ur home address & we’ll work it out. This will be kept private, only ur voting ward will be stored & u will be anonymous.” To determine each user’s polling station after the user entered their address required four-steps.<sup>12</sup> First, Vumi-Go invoked a script that dispatched the user input address to the Google API. Second, the Google API then matched the address. Third, if there was one match, then the user was matched to that address; if more than one match existed, the platform sent the top matches back to the user, who then had the option to choose one. If Google returned no match, the platform asked the user to re-enter her address and the process repeated. Fourth, the latitude and the longitude of the resulting Google API match placed the user within a polling station (for which we had pre-existing shape files from the Electoral Commission of South Africa) and the platform assigned the user to the polling station in which her coordinates were located. Fig. 3 displays matches to polling stations in Soweto Township.

Our implementing partners drove our decision to attempt to acquire this exact measure of user location (rather than directly asking

<sup>10</sup> The average district size is 1140 and there are 22,311 districts in the country. We designed our data collection to occur on this lowest level so that we could more likely detect an average treatment effect of our intervention.

<sup>11</sup> A note on how this may travel to other contexts: we think South Africa presents an interesting intermediate context in that it does not have the broad adoption of GPS-enabled smartphones, but does have a well-organized system of street addresses that makes the use of the Google API location lookup potentially feasible.

<sup>12</sup> We stored the address information separately from the rest of the record.



**Fig. 3.** The gray lines show the borders of each voting station. The shaded gray stations show stations where we successfully recruited individuals to monitor. The red line shows the border of the Soweto Township.

for the data we needed, the polling station); our partners believed that most people would not know the name of their polling station. Our secondary concern was that potential, but un-registered, voters with whom we engaged would have no polling station to report to us when we asked. Therefore, we attempted to have users share their address, from which we would back out their polling stations regardless of whether they had registered to vote.

Unfortunately, in practice, our original system failed to yield a true positive result over 90% of the time. Simply asking for an address drove attrition from the study (we believe this was both because the process was tedious, tended to time out and/or require multiple sessions in USSD, and because we did not make clear enough to users our purpose for collecting geolocation data). Given this high failure rate, the geographic lookup system failed to yield the number of subjects with a known polling station that we needed for our geographic study. Most users who entered any geographic information could not be matched to a polling station given the information they entered. Fundamentally, a majority of users (60%) did not enter any geographic information. Of the 53,622 users who registered before we turned off the geographic lookup, only 21,631 (40%) percent gave us *any* geographic information of quality. Of those, 5744 (26.4%) yielded any match in the Google API, representing just over 10% of the original sample. Hence, despite the 5744 matches that Google yielded (which likely include many false positives), this address lookup system fell short both in terms of data quality as well as in terms generating a sufficient absolute number of users for whom we had geographic data.

To understand better the types of problems our system generated, we randomly sampled 500 of the addresses where Google could not generate any match and carefully examined the usage patterns of these de-linked records. We developed a coding schema for three large categories of error types and a fourth residual category. The first category comprised incomplete information, defined as entries where individuals appear to have attempted to submit some quantity of the correct type of geographic

information, but not enough to produce a match. The second category defined instances where users entered the wrong type of information, for example a phone number, ward number, polling station name, or a P.O. Box. Users may also have submitted information that Google did not recognize. Our third classification involves extraneous information, including entries where individuals make no effort to enter useful information, such as expletives, expressions of their party support, or a few numbers. Using a simple set of regular expressions (a pattern recognition technique), we estimate approximately 42% of the errors come from incomplete information, 26% come from incorrect information, and 8% are extraneous information. The final 24% were in the residual category and difficult to classify.

Because of the poor quality of user-input geographic information, we reasoned that asking for geographic information caused subjects to not engage in the platform. To test this hypothesis, we re-contacted a subset of 10,000 users on USSD who had registered but not entered an address. In a geographic elicitation experiment, we randomly assigned these users to one of three groups. In the control group, users entered the system as in the original project conception. In the first treatment group (Reminder Treatment), the re-contact message included additional text reminding respondents to enter their geographic information. In the second treatment group (Geographic Lookup Off (GLO) Treatment), we turned off the geographic lookup. We then measure the intent to treat (ITT) effect of either reminding respondents about the geographic lookup or turning the system off by creating a binary variable where each observation received a 1 if they engaged in the most important section of dashboard that contained demographic questions (the Answer & Win (AW) section), and 0 otherwise. As a robustness check, we operationalize our dependent variable as a binary variable where users received a 1 if they engaged in any question in any part of the dashboard (ALL) beyond the geographic lookup, and 0 otherwise. Alternatively, we also operationalize our outcome as the count of questions answered in the entire platform or the count of questions answered in the first section.



**Table 5**  
Geographic lookup experiment.

	DV - AW		DV - ALL		DV Count -AW	DV Count All
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	0.08*** (0.01)	0.05* (0.02)	0.44*** (0.01)	0.41*** (0.04)	0.18* (0.09)	1.59*** (0.24)
GLO Treatment	0.11*** (0.01)	0.11*** (0.01)	0.05*** (0.01)	0.05*** (0.01)	0.41*** (0.03)	0.50*** (0.08)
Reminder Treatment	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.02 (0.03)	0.00 (0.08)
Days		0.00 (0.00)		0.00 (0.00)	0.01 (0.01)	0.03 (0.03)
Lottery		0.03** (0.01)		0.02 (0.01)	0.11** (0.04)	0.14 (0.10)
Subsidy		-0.01 (0.01)		-0.05** (0.02)	-0.03 (0.04)	-0.87*** (0.10)
R <sup>2</sup>	0.03	0.03	0.00	0.00	0.03	0.02
Adj. R <sup>2</sup>	0.03	0.03	0.00	0.00	0.03	0.02
Num. obs.	9999	9999	9999	9999	9999	9999
RMSE	0.32	0.32	0.50	0.50	1.20	3.33

In Model 1 and 2 the dependent variable is a binary variable for whether the respondent answered any of four key demographic questions. In Model 3 and 4 the dependent variable is a binary variable measuring if the respondent engaged in any part of the platform beyond the geographic lookup. In Model 5 and 6, we measure the count of key demographic questions the respondent answered out of a possible four (5) and a count of use of the system, as proxied by total questions answered in the entire system, with a theoretical maximum of 52 (6). Models 2, 4, 5 and 6 all contain control variables for the number of days the user was in the system at the time of the experiment and dummy variables for the different payment structures present on different USSD channels in the study \* p<0.1, \*\* p<0.05, \*\*\*p<0.01.

Table 5 models 1–4 show the regression coefficients for linear probability models. These regression demonstrate that asking for geographic information drives attrition, and that this finding is robust to controls – including days in the study and whether respondents had to pay for their participation in one of two ways (Lottery, Subsidy). In the experiment, we find that turning off the geographic lookup resulted in users being 11 percentage points more likely to re-engage in the platform compared to the original project conception as measured by answering key demographics questions (Model 1) or 5 percentage points more likely to answer any question (Model 3) among those that were re-contacted. Priming individuals about the necessity of the geographic lookup (Reminder Treatment), conversely, makes no difference in respondents’ engagement with our platform. The finding is also robust operationalization of the outcome variable, as a count as shown in Models 5 and 6. Since respondents were already re-contacted, this is likely an underestimate of the size of the attrition effect in the entire population, as many individuals, having already been contacted once, were likely to ignore our message and not dial back into the

**Table 6**  
Sequence of geographic questions.

Voting Station ID push	
MESSAGE	
2 b observer we need 2 know where u vote!	SCREEN 1
Tell us the Province u live in: 1. EC 2. FS 3. GP 4. KZN 5. Limp 6. MP 7. NCape 8 N West 9. WCape	
Thanks! Now we need ur town, suburb or district so we r sure of ur voting station	SCREEN 2
What is the name of your voting station? Pls type out. Example: KUNI PRIMARY SCHOOL	SCREEN 3
We'll use the IEC lookup service 2 confirm ur voting station with ur ID. Pls enter ID number OR Write "NO" & we'll try 2 ensure u can still observe if we verify ur voting station	SCREEN 4
Thanks 4 ur interest in observing the elections! Sadly we have enough ppl in ur location, but stay involved– ur voice is NB! U will get R5 for ur airtime use.	MESSAGE TO INCENTIVE GROUP 1
Thanks 4 volunteering 2 b an election observer! U will get R5 for airtime & R50 after u complete all tasks. U will get SMS about where u will b observing soon!	MESSAGE TO INCENTIVE GROUP 2

system. Such attrition could be particularly pernicious to representative estimates or population parameters if this selection is non-random, which is likely, since wealthier individuals clearly have more incentive to not disclose their address.

Due to the lack of true positive matches and the likelihood that asking for geographic information drove attrition in VIP:Voice, we turned off lookup permanently on April 17 (ten days after the system went live) – meaning that the platform did not ask any user who entered after this date to enter their address. This experience suggests that even in developing countries in which a usable database of street addresses exists (an extremely unusual condition in Sub-Saharan Africa), both willingness and data entry constraints are likely to make this kind of user-entered location lookup system an unattractive option.

Consequently, when we moved to Phase 3, the part of the study more directly focused on monitoring polling stations, we adjusted course and attempted to exploit an alternate means of acquiring location information. To improve on our earlier attempt at geospatial data collection, we conducted a focus group with young voters in the township of Vosloorus (outside of Johannesburg) to better understand how and what geospatial information they were willing to share and methods for doing so. This structured conversation revealed people did not feel uncomfortable sharing their polling station and asking this would not cause problems in the context of an election-related ICT platform. Conversely, focus group respondents felt strongly that they were unwilling to share their home address information.

We therefore set out to build an interactive process (working within the 140-character limitation of SMS) that would allow us to identify a user's polling station. Contrary to what our partners originally told us, we designed a set of sequential questions to undertake geospatial data collection in which we focused on obtaining users’ polling station information and did *not* ask for their home address. As shown in Table 6, our first question asks for vague geographic information while succinctly explaining to the user *why* we need such information. We then sequentially asked user for more fine-grained geospatial information. Using Vumi-Go's bi-directional system, after each of the first four screens, the user sent an SMS back to the system, which then triggered the next SMS. We built this system in SMS and not USSD because we were both worried about the timeouts and because we lacked the lead development time for USSD software development.

As shown in Table 6, to ensure that we accurately placed users in the correct polling station, we asked them to give us information about their geographic location *and* their voter identification number. We ran three scripts to check the quality of information. First, a script checked to determine whether each piece of geographic data was logically possible (e.g., the voting station was in the correct province). Second, if the user provided their voter id number, we entered this id into an automated service provided by the IEC to all South Africans; this service verified where the voter was registered to vote. Third, we compared the stated polling station with the district returned by the IEC lookup service. The highest quality match occurred when all declared geographic voting data were logically consistent and that data

matched the polling station returned by the IEC's service. In practice, given our statistical power calculations, we took all users who had the highest match or who entered geospatial information that met all our logical checks. We then enlisted these users as election monitors in our Phase 3 experimental protocol.

Using this new process, we were substantially more successful at obtaining robust data on users' voting location. While the context was different in that the task of monitoring elections was more onerous than signing up for a digital platform, of the 51,841 users of whom we invited to monitor, we received 3761 responses (7.2%). Of these responses, we could reliably identify the location of two-thirds (66.8%) of them, giving us 2512 eligible election monitors through our system.

We offer three important takeaways from these efforts to collect self-reported location data. First, asking for geographic information will likely drive attrition from the platform; therefore, researchers should not undertake it unless it is critical to their research design. Second, like in traditional Computer Assisted Telephonic Interview (CATI) survey work, researchers should solicit sensitive information such as geographic information as late in the data collection process as possible in interactive mobile and digital platforms after obtaining respondents' trust. Third, breaking down the solicitation of geographic information into a series of steps, each of which is engineered to be as robust as possible to errors in data entry, results in more accurate data. While asking sequential questions consumes more time and resources, it generates high quality data and allows for measurements of geographic data quality because researchers can employ the sequential nature of the data collection to test for non-logical response patterns.

### 3.3. Challenges to randomization

We explain the scope and protocol of our complex randomization protocols and discuss the challenges of implementing such protocols using Vumi-Go. We highlight that, while the process of randomization into groups (*assignment*) itself is not difficult, software application designers have not developed software that enables researchers to manipulate the sequence and timing of message *dispatch*.

Complex randomization in ICT platforms often have many treatment cells. The number of user groups for a randomization protocol that calls for blocking across  $c$  channels,  $t$  treatment groups, and  $j$  sub-treatments is the product:  $c*t*j$ . In our case, we had 2 channels, 3 treatment arms, and 90 sub-treatments, producing 540 user groups.

Randomization in any messaging (either bi- or uni-directional) system requires two different components for operationalization. The first, with which social scientists are familiar, is where researchers *assign* users to treatment and control groups. Assignment in statistical computing languages such as R is straightforward. Less well known to social science is the need for the ICT platform to *dispatch* the different messages assigned to users. A dispatcher — the computing engine that sends out the messages (since messages over SMS cannot all be sent at the same time, for example) — implements message *dispatch*. From a computer science perspective, the dispatcher must link the user to the group into which the researcher has randomized them, and then send the appropriate message.<sup>13</sup>

Vumi-Go fell short on both *assignment* and *dispatch*. Assignment within Vumi-Go required bespoke JavaScript code for each individual randomization. Since we were undertaking multiple randomizations, the development and testing cycle for each element did not allow for easy and quick turnaround. Vumi-Go also did not have an elegant way to assign many randomizations to dispatch, once we had randomized messages to users. In fact, the software architecture that Vumi-Go had developed for this — called *user groups* — was unwieldy and time consuming. It required the manual creation of “user groups” in Vumi

before platform users could be connected to those “user groups.” The dispatch mechanism, furthermore, was unable to randomize the timing of each message. While it is unlikely in our usage case that the exact time within an hour a message was sent was correlated with our outcome variables, for many applications, this might be the case.

We recommend researchers inquire carefully about whether an application's dispatch system can easily randomize both the assignment of individuals to different treatment groups, but, more importantly, easily link randomization groups to the dispatch of different messages, as well as randomize the timing of dispatch. We also hope that software engineers in the field of development engineering will consider these issues as they design their applications.

### 3.4. Cross-channel recruitment costs and representativity

The multi-channel implementation of this project places it in a unique position to offer comparative assessments of recruitment costs that we hope will be useful to researchers in understanding the relative benefits of different approaches to platform-level representativity.

We undertake a cost-effectiveness exercise whose purpose is to understand how much more it would have cost us to acquire additional users on each channel, and how this would have helped us to garner an overall platform population that more closely represents the national average derived from the South African census.<sup>14</sup> This analysis can be broken into a series of steps, represented in Table 7. First, to provide a simple way to think about representativity, Panel A presents the channel-specific demographic attributes (age, gender, and race) of each channel, and in the bottom row we present the national average. We see the overall platform usership was somewhat younger and had a higher proportion of Black South Africans than the national average, while the fraction of males in our sample almost exactly matches the country as a whole (51%). The only channel yielding users who are both older and more likely to be White or Asian than the national average is Twitter/Gtalk, immediately indicating that it is increasing usership of social media channels that will likely make the platform more nationally representative.

To operationalize the cost of achieving national representativity, Panel B breaks down all the costs of the platform into per-channel fixed costs required to field the platform at all (column A), and per-channel marginal costs that were incurred in advertising or outreach and whose fundamental purpose is to increase the number of users (column B). Some costs are spread across multiple channels, or across the whole platform; we apportion these fixed multi-channel costs to specific channels proportional to usage.

Using these inputs (Panel B, columns A-C), we estimate the cost per additional user for each channel, which is the amount that we spent to acquire users on that channel divided by the number of users we obtained (panel B, column E). Assuming that these costs would continue to scale linearly on the margin, this gives us what it would take to obtain one more user on each channel (panel B, column F). These costs vary widely, from a low of \$2 on Mxit to \$250 and up for Twitter and Mobi.

To gauge how additional users would affect the representativity of the platform, we then calculate the “representativity benefit” of each channel by asking what would happen to the platform average outcome if we added 5000 users to that channel (this is roughly the average number of registered users per channel on the platform, and so represents a doubling for the average channel). This is done using a logic similar to the Mahalanobis weighting criterion, where we calculate the absolute deviation of the platform average from the national average for each demographic attribute, divide this by the

<sup>13</sup> In a theoretical world, the dispatcher would also randomize the order/timing of the dispatch of each message.

<sup>14</sup> A related paper on VIP:Voice (Ferree et al., 2017) conducts a very different type of cost benefit analysis, considering the primary goal of the platform as being to recruit citizen election monitors, and comparing the costs over our system to alternate monitoring technologies and modalities.

**Table 7**  
Cost effectiveness.

<b>Panel A: Representativity:</b>						
	Demographic Characteristics by Channel:					
	A.	B.	C.	D.	E.	F.
Channel Average:	Age	Male	Black	Coloured	White	Asian
USSD Lottery	26.4	0.353	0.936	0.050	0.009	0.005
USSD Free	25.9	0.343	0.938	0.044	0.009	0.009
USSD Standard	25.1	0.352	0.942	0.047	0.008	0.002
USSD Other/Newspaper	26.8	0.348	0.939	0.046	0.015	0.001
Mxit	22.8	0.622	0.816	0.148	0.023	0.013
Mobi	23.7	0.350	0.890	0.087	0.015	0.007
Twitter/Gtalk	25.5	0.485	0.639	0.115	0.131	0.115
<b>Platform Average</b>	24.1	0.517	0.862	0.110	0.018	0.010
<b>Census National Average</b>	24.9	0.510	0.792	0.094	0.089	0.025

<b>Panel B: Cost versus Representativity:</b>								
Channel:	A.	B.	C.	D.	E.	F.	Representativity benefit of an additional 5000 users on this channel	Representativity benefit of \$50,000 extra dollars on this channel
	Channel-Specific Fixed Costs	Channel-Specific Average Variable costs (to acquire more users)	Total spending per channel, actual	Number registered users who provided demographic info	Average Total Cost per User (C/D)	Average Variable Cost per User (B/D)		
USSD Lottery	\$30,023	\$69,856	\$99,879	7415	\$13.47	\$9.42	-0.014	-0.016
USSD Free	\$9483	\$14,088	\$23,571	2342	\$10.06	\$6.02	-0.022	-0.049
USSD Standard	\$6401	\$8498	\$14,899	1581	\$9.42	\$5.38	-0.045	-0.095
USSD Other/Newspaper	\$6132	\$6547	\$12,678	1218	\$10.41	\$5.38	-0.006	-0.029
Mxit	\$58,715	\$41,661	\$100,376	20,078	\$5.00	\$2.07	-0.027	-0.086
Mobi	\$18,603	\$20,318	\$38,921	64	\$608.14	\$317.47	-0.038	0.001
Twitter/Gtalk	\$5250	\$16,571	\$21,822	67	\$325.70	\$247.34	0.370	0.017
<b>Totals</b>	<b>\$134,606</b>	<b>\$177,540</b>	<b>\$312,146</b>	<b>32,765</b>				

standard deviation of that outcome in the population, and take the unweighted sum of these normalized deviations as a goodness of fit measure. In other words, the representativity score of the platform is calculated as  $\sum_i \frac{abs(x_i - x_{iN})}{\sigma_i}$ , where  $x_i$  is the platform average for demographic indicator  $i$ ,  $x_{iN}$  is the national average for that indicator, and  $\sigma_i$  is the standard deviation of that indicator. By calculating this score before and after adding additional users to each channel we can simply summarize the effect of increasing the size of a channel on the platform-level representativity. A positive number in the ‘Representativity benefit’ column indicates that the increase in users of that type on that channel pushes the platform average towards the national average. The final column incorporates the cost per additional user calculated in Column F to ask what would happen had we spent an additional \$50,000 on each channel. Here we see that only Twitter/Gtalk, given their high proportions of older and whiter users, has a positive effect on the representativity of the national platform.

This exercise illustrates the cost versus representativity trade-off: those channels that were most cost effective typically had very large numbers of users and hence the platform average is weighted towards them. In our context, we achieved high success rates with two technologies: USSD and Mxit, both of which are, or were, ubiquitous in South Africa. Channels such as Mobi, Twitter, and Gtalk have yet to achieve the enormous user bases there that they have elsewhere in the world, and hence despite the fact that they garnered the older and whiter users needed to achieve representativity, they were less cost effective per user. The broad lesson is that to achieve breadth it may be necessary to invest in technology channels that are relatively less cost effective, because they are intrinsically likely to be the ones used by the people underrepresented on the platform.

#### 4. Conclusions: lessons learned and advice moving forward

This paper describes a novel cross-channel ICT platform deployed to study citizen engagement and election monitoring in South Africa’s 2014 national elections. Our results provide social science researchers looking to scale their intervention and applied technologists aiming to tailor their solutions to researcher specifications important insights

into multi-channel platform design. We document our platform’s advantages over traditional modes of data collection, including the low marginal costs of additional users, better ability to obtain demographic representativeness and the ability to collect rich data on the timing of response. We also document challenges of such multi-platform systems including eliciting geographic information, ensuring comparability across platforms, and properly randomizing users.

An important caveat to our results is that, due to the nature of an election-related online platform, we stood up a system over a relatively short period and only solicited interaction over a single month. While this is typical of event-driven online platforms, it may not represent well the broader set of ICT platforms whose purpose is to build a durable, longer-term user base (such as for health reporting, market interactions, and e-governance more broadly). In this sense, our results should be read as informing implementers who are seeking to stand up national-scale systems quickly, but may not provide as much guidance, or on all aspects, for projects whose goals are to build long-term relationships with their user bases.

We share three areas where further research on cross-channel ICT platforms should occur. First, collecting geospatial data in a manner that comports with research ethics and is voluntary requires additional study and development. Second, researchers need to pay more attention to understand how to deploy shorthand versions of questions in multiple languages across different technologies. Third, development engineering platforms should build randomization techniques into the heart of their infrastructure. Moving forward, we recommend close collaboration between social and computer scientists to build and field technologies that coincide with the cutting-edge scholarship in the social and behavioral sciences. We see tremendous promise in studying how purely digital platforms such as our own can interact with more traditional forms of engagement to bolster participation and facilitate high quality data collection even further.

#### Acknowledgements

We thank Matthew DeGale, Pippa Yeats, Lieze Langford, and colleagues at the Praekelt Foundation for their help building the platform. We also thank Livy and Code4SA. Kate Blackwell, Grant

Buckles, and Alex Verink provided excellent research assistance. All mistakes remain with the authors and any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of USAID or the Praekelt Foundation.

**Funding**

This work was supported by the U.S. Agency for International Development (USAID), Development Innovation Ventures (AID-OAA-A-14-00004) and the Harvard Academy for International and Area Studies [Long].

**Appendix**

*Phase 2 Description*

In Phase 2, Participants continued engagement through their enrollment channel. The “What’s Up?” survey asked questions on local campaign activities, while “VIP” posed relatively standard polling questions on participation in local events, evaluation of ANC performance, and probability of voting.

In addition to these surveys, presented via drop-down menus, VIP:Voice tracked real-time shifts in political opinion and incidents of political activities. One set of questions, the “Activity” survey, asked about local political activities at three different times prior to election day, randomizing the day on which an individual received the survey A second set of “Thermometer,” questions asked about voting intentions and party support. Users could complete surveys in any order, and failure to complete one survey did not preclude answering questions on others. Phase 2 required digital forms of engagement as all activities were limited to interacting with the platform. Of the 90,646 people registered, 34,727 (38%) completed the four demographic questions and 15,461 (17%) answered the demographic questions and one of the other four Phase 2 surveys. See also [Ferree et al. \(2017\)](#) for additional details.



**Fig. A1.** Example splash screens for recruitment into platform.

**Table A1**  
Descriptive statistics.

Variable	Mean	Standard Deviation	Min	Max
All Questions (#)	1.86	3.35	0	37
Answer & Win Questions (#)	0.44	1.22	0	4
Control	0.14	0.35	0	1
Days in System	6.83	1.53	3	10
Engaged in System	0.12	0.33	0	1
Lottery Treatment	0.58	0.49	0	1
No Geographic Lookup	0.33	0.47	0	1
Normal Geographic Lookup	0.33	0.47	0	1
Primed Geographic Lookup	0.33	0.47	0	1
Subsidy Treatment	0.18	0.38	0	1

## References

- Aker, J.C., 2011. Dial “A” for agriculture: a review of information and communication technologies for agricultural extension in developing countries. *Agric. Econ.* 42, 631–647. <http://dx.doi.org/10.1111/j.1574-0862.2011.00545.x>.
- Aker, J.C., Ksoll, C., Lybbert, T.J., 2012. Can mobile phones improve learning? evidence from a field experiment in Niger. *Am. Econ. J. Appl. Econ.* 4, 94–120. <http://dx.doi.org/10.1257/app.4.4.94>.
- Bailard, C.S., 2012. A field experiment on the internet's effect in an African election: savvy citizens, disaffected voters, or both? *J. Commun.* 62, 330–344. <http://dx.doi.org/10.1111/j.1460-2466.2012.01632.x>.
- Bailard, C.S., Livingston, S., 2014. Crowdsourcing accountability in a Nigerian election. *J. Inf. Technol. Polit.* 11, 349–367. <http://dx.doi.org/10.1080/19331681.2014.947056>.
- Benoit, K., Conway, D., Lauderdale, B.E., Laver, M., Mikhaylov, S., 2016. Crowd-sourced Text Analysis: Reproducible and Agile Production of Political Data. *Am. Political Sci. Rev.* 110, 278–295. <http://dx.doi.org/10.1017/S0003055416000058>.
- Blumenstock, J., Cadamuro, G., On, R., 2015a. Predicting poverty and wealth from mobile phone metadata. *Science* 350, 1073–1076. <http://dx.doi.org/10.1126/science.aac4420>.
- Blumenstock, J., Callen, M., Ghani, T., Koepke, L., 2015b. Promises and Pitfalls of Mobile Money in Afghanistan: Evidence from a Randomized Control Trial. In: *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development, ICTD '15*. ACM, New York, NY, USA, p. 15:1–15:10. <https://doi.org/10.1145/2737856.2738031>.
- Brunette, W., Sundt, M., Dell, N., Chaudhri, R., Breit, N., Borriello, G., 2013. Open data kit 2.0: expanding and refining information services for developing regions. In: *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*. ACM, p. 10.
- Bussell, J.L., 2010. Why get technical? Corruption and the politics of public service reform in the Indian states. *Comp. Polit. Stud.* 43, 1230–1257. <http://dx.doi.org/10.1177/0010414010369073>.
- Callen, M., Gibson, C.C., Jung, D.F., Long, J.D., 2016. Improving electoral integrity with information and communications technology. *J. Exp. Polit. Sci.* 3, 4–17. <http://dx.doi.org/10.1017/XPS.2015.14>.
- Chen, J.V., Ross, W., Huang, S.F., 2008. Privacy, trust, and justice considerations for location-based mobile telecommunication services. *info* 10, 30–45. <http://dx.doi.org/10.1108/14636690810887526>.
- Chicoine, L., Guzman, J.C., 2017. Increasing Rural Health Clinic Utilization with SMS Messages: Evidence from a Randomized Evaluation in Uganda. *World Dev.* 99, 419–430. <https://doi.org/10.1016/j.worlddev.2017.05.029>.
- Dhaliwal, I., Hanna, R., 2017. The devil is in the details: the successes and limitations of bureaucratic reform in India. *J. Dev. Econ.* 124, 1–21. <http://dx.doi.org/10.1016/j.jdevco.2016.08.008>.
- Dillman, D.A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., Messer, B.L., 2009. Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Soc. Sci. Res.* 38, 1–18. <http://dx.doi.org/10.1016/j.ssresearch.2008.03.007>.
- DiMaggio, P., Hargittai, E., Neuman, W.R., Robinson, J.P., 2001. Social implications of the internet. *Annu. Rev. Sociol.*, 307–336. <http://dx.doi.org/10.1146/annurev.soc.27.1.307>.
- Ferree, K.E., Gibson, C.C., Jung, D.F., Long, J.D., McIntosh, D., 2017. How Technology Shapes the Crowd: Participation in the 2014 South African (Working Paper No. 067), CEGA Working Paper Series. Center for Effective Global Action, University of California, Berkeley. <https://escholarship.org/uc/item/5pd6k715>.
- Findley, M.G., Nielson, D.L., Sharman, J.C., 2013. Using field experiments in international relations: a randomized study of anonymous incorporation. *Int. Organ.* 67, 657–693. <http://dx.doi.org/10.1017/S0020818313000271>.
- Grossman, G., Humphreys, M., Sacramone-Lutz, G., 2014. “I wld like u WMP to extend electricity 2 our village”: on information technology and interest articulation. *Am. Polit. Sci. Rev.* 108, 688–705. <http://dx.doi.org/10.1017/S0003055414000331>.
- Hardy, H., Kumar, V., Doros, G., Farmer, E., Drainoni, M.-L., Rybin, D., Myung, D., Jackson, J., Backman, E., Stanic, A., Skolnik, P.R., 2011. Randomized controlled trial of a personalized cellular phone reminder system to enhance adherence to antiretroviral therapy. *AIDS Patient Care STDs* 25, 153–161. <http://dx.doi.org/10.1089/apc.2010.0006>.
- Hartung, C., Lerer, A., Anokwa, Y., Tseng, C., Brunette, W., Borriello, G., 2010. Open Data Kit: Tools to Build Information Services for Developing Regions. In: *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*. ACM, p. 18.
- Heeks, R., 2008. ICT4D 2.0: the next phase of applying ICT for international development. *Computer* 41, 26–33. <http://dx.doi.org/10.1109/MC.2008.192>.
- Humphreys, M., Weinstein, J., 2012. *Policing Politicians: Citizen Empowerment and Political Accountability in Uganda Preliminary Analysis*. Columbia Stanf. Univ. Unpubl. Manusc.
- Källander, K., Tibenderana, J.K., Akpogheneta, O.J., Strachan, D.L., Hill, Z., Asbroek, A.H.A., ten, Conteh, L., Kirkwood, B.R., Meek, S.R., 2013. Mobile Health (mHealth) Approaches and Lessons for Increased Performance and Retention of Community Health Workers in Low- and Middle-Income Countries: a review. *J. Med. Internet Res.* 15, e17. <http://dx.doi.org/10.2196/jmir.2130>.
- Kalton, G., Schuman, H., 1982. The effect of the question on survey responses: a review. *J. R. Stat. Soc. Ser. Gen.* 145, 42–73. <http://dx.doi.org/10.2307/2981421>.
- Kuntsche, E., Labhart, F., 2013. ICAT: development of an internet-based data collection method for ecological momentary assessment using personal cell phones. *Eur. J. Psychol. Assess.* 29, 140–148. <http://dx.doi.org/10.1027/1015-5759/a000137>.
- Kuntsche, E., Robert, B., 2009. Short Message Service (SMS) technology in alcohol research—a feasibility study. *Alcohol Alcohol.* 44, 423–428. <http://dx.doi.org/10.1093/alcalc/agg033>.
- Lim, M.S.C., Hocking, J.S., Hellard, M.E., Aitken, C.K., 2008. SMS STI: a review of the uses of mobile phone text messaging in sexual health. *Int. J. STD AIDS* 19, 287–290. <http://dx.doi.org/10.1258/ijsa.2007.007264>.
- Madon, S., Reinhard, N., Roode, D., Walsham, G., 2009. Digital inclusion projects in developing countries: processes of institutionalization. *Inf. Technol. Dev.* 15, 95–107. <http://dx.doi.org/10.1002/itdj.20108>.
- McEwan, P.J., 2015. Improving learning in primary schools of developing countries a meta-analysis of randomized experiments. *Rev. Educ. Res.* 85, 353–394. <https://doi.org/10.3102/0034654314553127>.
- Merrill, J., Hershov, R., Gannett, K., Barkley, C., 2013. Pretesting an mHealth Intervention for At-risk Adolescent Girls in Soweto, South Africa: Studying the Additive Effects of SMSs on Improving Sexual Reproductive Health & Rights Outcomes. In: *Proceedings of the Sixth International Conference on Information and Communications Technologies and Development: Notes - Volume 2, ICTD '13*. ACM, New York, NY, USA, pp. 96–99. <https://doi.org/10.1145/2517899.2517933>.
- Okolloh, O., 2009. Ushahidi, or testimony: Web 2.0 tools for crowdsourcing crisis information. *Particip. Learn. Action* 59, 65–70.
- Paluck, E.L., Blair, G., Vexler, D., 2010. *Entertaining, informing, and discussing: behavioral effects of a democracy-building radio intervention in Southern Sudan*. Univ. Manusc. Prep. Princet. Univ.
- Pop-Eleches, C., Thirumurthy, H., Habyarimana, J.P., Zivin, J.G., Goldstein, M.P., de Walque, D., MacKeen, L., Haberer, J., Kimaiyo, S., Sidle, J., Ngare, D., Bangsberg, D.R., 2011. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. *AIDS* 25, 825–834. <http://dx.doi.org/10.1097/QAD.0b013e32834380c1>.
- Rajput, Z.A., Mbugua, S., Amadi, D., Chepngeno, V., Saleem, J.J., Anokwa, Y., Hartung, C., Borriello, G., Mamlin, B.W., Ndege, S.K., Were, M.C., 2012. Evaluation of an Android-based mHealth system for population surveillance in developing countries. *J. Am. Med. Inform. Assoc.* 19, 655–659. <http://dx.doi.org/10.1136/amiainj-2011-000476>.
- Rezaee, A., Hasanain, A., Khan, Y., 2015. *Crowdsourcing government accountability: Experimental evidence from Pakistan* (CEG Working Paper No. 123). Global Economic Governance Programme, Oxford.
- Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B., Rand, D.G., 2014. Structural topic models for open-ended survey responses. *Am. J. Polit. Sci.* 58, 1064–1082. <http://dx.doi.org/10.1111/ajps.12103>.
- Sachs, J.D., McArthur, J.W., 2005. The millennium project: a plan for meeting the millennium development goals. *The Lancet*. 365, 347–353.
- Schuman, H., 1996. *Questions and Answers in Attitude Surveys: Experiments on Question form, Wording, and Context*. Sage Publications, Thousand Oaks, CA.
- Shaikh, A.A., Karjaluoto, H., 2015. Mobile banking adoption: a literature review. *Telemat. Inform.* 32, 129–142. <http://dx.doi.org/10.1016/j.tele.2014.05.003>.
- Shilton, K., 2009. Four billion little brothers?: privacy, mobile phones, and ubiquitous data collection. *Commun. ACM* 52, 48. <http://dx.doi.org/10.1145/1592761.1592778>.
- Swendeman, D., Rotheram-Borus, M.J., 2010. Innovation in sexually transmitted disease and HIV prevention: internet and mobile phone delivery vehicles for global diffusion. *Curr. Opin. Psychiatry* 23, 139–144. <http://dx.doi.org/10.1097/YCO.0b013e328336656a>.
- Thompson, M., 2008. ICT and development studies: towards development 2.0. *J. Int. Dev.* 20, 821–835. <http://dx.doi.org/10.1002/jid.1498>.
- Tomlinson, M., Rotheram-Borus, M.J., Swartz, L., Tsai, A.C., 2013. Scaling up mHealth: where is the evidence? *PLOS Med.* 10, e1001382. <http://dx.doi.org/10.1371/journal.pmed.1001382>.
- Walsham, G., Sahay, S., 2006. Research on information systems in developing countries: current landscape and future prospects. *Inf. Technol. Dev.* 12, 7–24. <http://dx.doi.org/10.1002/itdj.20020>.
- Warschauer, M., 2003. *Technology and Social Inclusion: Rethinking the Digital Divide*. MIT Press, Cambridge, Mass.
- Wouters, B., Barjis, J., Maponya, G., Martiz, J., Mashiri, M., 2009. Supporting home based health care in South African rural communities using USSD technology. *AMCIS 2009 Proceedings* 410.