Mobile Phones for Health Education in the Developing World: SMS as a User Interface

Catalina M. Danis,[1] Jason B. Ellis,[1] Wendy A. Kellogg,[1] Hajo van Beijma,[2] Bas Hoefman,[3] Steven D. Daniels,[1] Jan-Willem Loggers [4]

[1]IBM T.J. Watson Research Center, P.O. Box 704, Yorktown Heights, NY 10598 USA
[2]Text to Change Netherlands, Reinaert de Vosstraat 13, Amsterdam, 1055 CL The Netherlands
[3]Text to Change Uganda, Oroma Villa, Plot#83, Katalina Road (OPP Kembabazi Catering Centre) Naguru, P.O. Box No. 21428, Kampala, Uganga
[4]Text to Change Uganda, Bukoto Crescent, Plot nr. 31 Naguru P.O. Box: 24134, Kampala, Uganda

{danis l jasone l wkellogg l danields @ us.ibm.com} {hvanbeijma l bhoefman l jwloggers @ texttochange.com}

ABSTRACT

Uganda suffers from a severe shortage of professional healthcare workers. Thus, programs aimed at prevention of disease are an important complement to the limited healthcare delivery system. We analyze two deployments of an SMS-based HIV/AIDS education system that uses a quiz format to assess people's knowledge of the disease, including its causes and methods of prevention. The deployments were to two groups in Uganda, one a sample of mobile phone users who live in a town in northwest Uganda; the other, workers at three factories in central and southeastern Uganda. The two samples had accuracy rates above chance levels and workers at the three factories had higher rates of participation (more individuals and more questions) than the sample selected from the cell tower service area. An analysis of incorrect answers suggested that while participants had some difficulty in matching the formatting required by the quiz, literacy did not appear to be a significant problem. We discuss the results in terms of implications for using SMS as a user interface mechanism and explore the possibility of using social ties among participants as a way to promote the scalability and sustainability of this quiz-based education method.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: H.4.3. Communications Applications.

General Terms

Measurement, Design, Experimentation, Human Factors.

Keywords

Text to Change, HIV/AIDS awareness, mHealth, ICT4D, developing nations, social computing, SMS.

ACM DEV'10, December 17-18, 2010, London, United Kingdom.

1. INTRODUCTION

There is an average ratio of one doctor for every 20,000 people in Sub-Saharan Africa [1] and the prevalence rate for diseases such as HIV/AIDS remains high [2]. Due to these human resource and disease challenges, there is a pressing need for new health approaches that are focused on prevention rather than treatment and that do not depend on traditional face-to-face interactions between medical providers and the public.

This paper discusses the evaluation of a mobile phone application designed to address some of these challenges. The system we studied leverages the broad penetration of mobile phones in Africa to deliver HIV/AIDS education through incentive-based Short Message Service (SMS) quizzes. The goal of deploying HIV/AIDS related quizzes to the general population is to increase awareness and knowledge about significant public health concerns, to promote healthy behaviors, and to encourage people to know their health status and seek treatment if necessary. The system was deployed by a non-governmental organization (NGO) from the Netherlands called Text to Change (TTC) that works in partnership with healthcare NGOs in Africa to develop appropriate medical content for the quizzes.

We present results from our analysis of usage data from two deployments: 1) a group of 10,000 mobile phone users in a circumscribed geographic area of northwest Uganda, and 2) a total of almost 5,000 workers in three factories located in central and southeastern Uganda. Our research focuses on two questions. First, we investigate whether interaction over SMS presents an effective user interaction mechanism for these populations of users. SMS is an attractive choice for a user interface (UI) given the ubiquity of SMS-capable phones in Africa but the question remains whether the general public can use the modality effectively. Second, we explore factors that influence participation in these sorts of quizzes. We are particularly interested in determining whether social factors might take the place of the often-used monetary incentives to drive participation rates.

The remainder of this paper is organized as follows. We first provide some background on healthcare challenges in Sub-Saharan Africa that may be addressable through information and communications technology (ICT) and characterize the state of mobile phone penetration on the continent. Next, we review related work on "mHealth" applications, with a specific focus on the use of SMS as a mode of interaction by both trained health workers and the general public. We then describe the TTC program for HIV/AIDS awareness and education that utilizes mobile phone-based quizzes targeted at rural participants. The majority of the paper details the two deployments of the TTC quiz system in rural Uganda and our analysis to evaluate the efficacy of the UI based on SMS and the rates of participation in the program. We conclude with a discussion of the findings and their implication for SMS-based quiz applications.

2. BACKGROUND

2.1 Healthcare Challenges

The healthcare system in Sub-Saharan Africa is under significant strain. While the WHO recommends a 1:1,000 ratio of doctors to patients, the average ratio of doctors per patient in Africa is 1:20,000 and this ratio can be as low as 1:50,000 in countries like Malawi and Tanzania [1]. The ratio for Uganda is 1:10,000; however, 70% of doctors and 40% of the nurses are concentrated in the urban areas (home to 12% of the total population), leaving many fewer resources for rural areas [3]. In Sub-Saharan Africa, there were 22.4 million people living with AIDS as of 2008. During that same year, 1.4 million people died from the disease and 1.9 million were newly infected (of which nearly 400,000 were children). These numbers have not changed significantly since 2001 [2] though they have dropped from the peak levels in 1996.

In order to address the scarcity of trained health professionals in Sub-Saharan Africa, many countries have deployed networks of lay health workers with basic training as the first line of primary care. Such workers take pressure off of the health system by performing basic services in the community where they live. Duties can include supporting the maintenance of treatment regimens for HIV/AIDS and tuberculosis, monitoring household health, basic differential diagnosis (e.g., malaria vs. pneumonia), and sharing information with those above them in the healthcare hierarchy [4].

2.2 Mobile Phone Use in Africa

Another way to augment strained health systems is through the appropriate application of ICTs – in this case, mobile phones. The dramatic penetration of mobile phones in Africa has set the stage for tremendous innovation in mobile applications and services. The number of mobile phone users in Africa grew from 49 million in 2002 to 280 million in 2007 and is projected to approach 600 million by 2012 [5]. This widespread adoption in a continent of approximately one billion people, combined with the relative dearth of wired network infrastructure and computers means that many Africans' first contact with a networked computing device is the mobile phone. This is particularly true of the rural poor for whom "pay as you go" mobile service provides access to technology that would otherwise be unaffordable. It is the value that mobile phones provide – not only as a personal communication medium but also as a way to improve one's livelihood (e.g., fishermen who use their mobiles to find a good place to sell their fish, enabling better profits while simultaneously driving down the cost for local consumers [6]) – that causes the very poor to make sacrifices in order to have a phone. Numerous practices have evolved that allow for cost-effective use of basic phones, from use of "beeping" or intentionally missed calls to village "phone ladies" who enable distant relatives to send remittances to loved ones using airtime as currency [7, 8].

3. RELATED WORK

Numerous applications have explored the possibility of addressing health challenges using mobile phones and other mobile devices in Africa and other regions of the developing world. These applications, referred to as mobile health (mHealth), are critical in places where existing health infrastructure cannot meet demand. A 2009 United Nations/Vodafone Foundation study found that for the 5.3 billion people living in developing countries, there are 2.3 billion mobile phone users, but only 305 million computers and 11 million hospital beds. Thus, the report concludes, "mobile phones reach further into developing countries than other technology and health infrastructure" [9].

mHealth applications fall into five broad categories: remote data collection, remote monitoring, communication and training for healthcare workers, diagnostic treatment support, and education and awareness [9]. A key early application, EpiHandy, [10] provided form-based data entry capability on handheld computers (PDAs like PalmPilot). It solved some key problems with traditional paper-based solutions, including error reduction methods such as in-the-field data correctness checks and the elimination of handwritten forms, which often take significant time to reach the appropriate ministries and then must be input into computers (an error-prone process).

Early applications were developed to run on PDAs, but many have transitioned to mobile phones as available phones have become more capable. There are at least two reasons for this transition. First, while mobile phones need to be re-charged just like PDAs, researchers have observed that workers with limited access to power are more likely to recharge their mobile phones than their PDAs. Hence, deploying a single device capable of supporting both data and personal communication is a better choice. Second, PDAs do not typically have internet access. Therefore, users must wait until they return to a location with network connectivity to transmit their data to a central server for further analysis (sometimes delaying the receipt of that data by a significant period of time).

One widely used mobile phone forms-based platform is the open source, community-developed OpenXdata platform [11].

3.1 SMS-based Applications

SMS-based mobile applications are an important class of mHealth applications because even the low-end "brick" mobile phones widely deployed among the African poor support SMS as an interaction modality. While the popular mode of use for mobile phones can vary significantly by region[1], SMS predominates in much of Africa [13]. The development of SMS-based mHealth applications is facilitated by the availability of application frameworks such as FrontlineSMS [14] and RapidSMS [15] that enable the sending and receiving of group SMS messages. One of the many applications developed on top of RapidSMS is ChildCount [16], with which community health workers in Kenya were trained to monitor malnutrition and malaria in children under the age of five. Health care workers communicated results through structured SMS messages (messages composed of a series of keywords followed by data values) following a two to three-hour training session that covered the use of the mobile phones and the

[1] For example, in South Africa the low cost of sending data via mobile phone has created a situation where users prefer to interact with one another online [12] rather than through SMS or voice.

correct formatting of the SMS messages. Even with training about 10% of messages were rejected by the system because of improper formatting and additional training was required whenever new fields (for the collection of new measurements) were introduced. Another use of structured SMS messages by trained workers was SMS BloodBank [17], in which nurses were trained to send SMS messages to request replenishment of blood products from a central location. The main obstacle in that deployment was the cost of the SMS message to the nurses. This very efficient mechanism was in danger of failing until the developers reimbursed nurses for the sending of messages. ELMR is an SMS-based application for updating and querying lightweight mobile health records [18]. Intended for use not only by health care workers but also by patients (though perhaps with more limited functionality), the intent is to allow users to interact with a database table through SMS.

SMS applications are also widely used by the general public, but since training is generally not possible, there is a trade-off between placing fewer requirements on the user (e.g., such as accepting free-form text) and requiring the computer receiving the messages to do more work to extract meaning from the input. This is certainly the challenge in the widely used Ushahidi crisis crowd-sourcing application [19] that accepts unstructured information from the public through various modalities, including SMS, and uses complex back-end software (SwiftRiver) to validate and filter the crowd-sourced information. Another project, called Tweak the Tweet, from the EPIC group at the University of Colorado at Boulder attempts to get the general public to structure their messages without the benefit of organized training by socializing the formatting conventions through viral means. This group of trained Twitter users "retweets" unstructured messages in the proper format, creating new messages that are machine-readable [20].

3.2 Alternatives to SMS-based Applications

Written literacy is a key challenge for IT systems deployed in developing nations; hence, many application developers have elected to use speech-based interaction as an alternative to SMS. For example, HealthLine [21], an education application for community health workers in Pakistan provides information via voice in the local language. Callers speak the health topic they are interested in learning more about and the system provides a set of voice menus that allows them to navigate to the information they need. Spoken Web [22], a platform that enables the creation and navigation of user-generated content using speech interaction, has been deployed, primarily in India, to a variety of user groups, including farmers and micro-business owners. Other applications incorporate automatic speech recognition technology (ASR). However, there are serious challenges in deploying ASR systems in the developing world, including the need to support multiple dialects and/or languages, noisy use conditions, and poor quality phone lines [23]. In addition, the typical pay-as you-go mobile phone plans do not provide "voice mail" thus effectively preventing asynchronous communication.

In summary, a variety of UI modalities have been applied in the service of mHealth – from SMS through ASR technology running on anything from basic phones through smartphones, and PDAs. Each modality has its challenges that must be addressed through appropriate design in the context of the particular user group and application needs.

4. AN SMS QUIZ APPLICATION

The Dutch and Uganda-based NGO Text to Change developed the mobile phone application that we discuss in this paper. They were founded in 2006 with the overall goal of providing health education in the developing world via mobile phones. More specifically, TTC has four primary goals: 1) to gauge the knowledge of a population about HIV/AIDS in order to inform the content of future quizzes and better target interventions, 2) to educate participants about safe behaviors with regard to HIV/AIDS, 3) to motivate people to seek HIV testing and treatment, and 4) to reduce HIV/AIDS-related stigma and discrimination.

The approach used in the TTC application can be summarized as an incentive-based mobile phone quiz: questions are sent to the participants' mobile phones, participants reply, and those participants with the highest accuracy and participation rates are entered into drawings for free "airtime" (pre-paid mobile phone minutes) and other prizes (see Figure 1). The quiz format is intended to provide a fun way for participants to engage with educational content.

1. TTC System sends a question to participants
2. Participants answer the question
3. TTC System checks the answer and responds:
    a) If answer is correct, participants get more information on the topic
    b) If answer is wrong, participants get the correct answer and an explanation
4. Repeat from step 1 until quiz is complete

Figure 1: TTC interaction model

Participants interact with the quiz via SMS (text messaging), and SMS presents at least two challenges when used as a user interface. Firstly, each message must conform to the 160-character limit (a technical limit imposed by the format). Secondly, the format (text) requires that the user be literate in the language of the quiz. We discuss the impact of each of these challenges below in the context of the deployments we studied.

The deployment of a quiz requires cooperation among a number of stakeholders. TTC typically works with a partner organization that has a specific population in mind for participation and has the necessary relationships to provide access to the population. In addition, they require cooperation from telecommunication service providers. TTC deployed its first mobile quiz in 2008 and to date has completed nine quiz deployments, reaching roughly 60,000 participants in Uganda, Kenya, and Namibia. Targeted populations most commonly consist of the rural poor.

5. METHOD

5.1 Deployment Sites

The analyses presented in this paper are of data collected in two quiz deployments by TTC and their NGO partners. The deployments differed on details of participant selection, duration of the quiz period, and the number and type of questions asked.

The first deployment, referred to as the *DistrictQuiz* deployment, took place from 28 January 2009 through 6 April 2009 in a rural district in northwest Uganda. TTC worked in conjunction with a local telecommunications provider to identify a set of 10,000 mobile phone numbers that were highly active within the district during a two-week period. TTC then sent messages to these phones to invite their "owners" to participate in the quiz[2]. In addition to targeted invitations to participate in the quiz, the *DistrictQuiz* was publicized through various media outlets, principally newspapers and radio stations serving the district. It is not known how effective this method was at creating awareness of the quiz deployment among those targeted for participation. TTC's healthcare NGO partner created the multiple-choice 13-question HIV/AIDS quiz.

The second deployment, called *FactoryQuiz*, was comprised of three parallel deployments to workers in three factories in central and southeastern rural Uganda. These three samples will be referred to as *Factory1*, *Factory2*, and *Factory3*. Two of the factories process sugar, while the third mines cobalt. The factories are located within factory towns that provide housing, schools, and health facilities to the employees and their families.

The quizzes were deployed to the factory workers from 13 August 2009 through 30 September 2009 and consisted of 18 true/false questions created by a different healthcare NGO partner from the first deployment. Workers were recruited to participate through face-to-face interviews with specially trained HIV/AIDS "peer educators" who are part of each factory's education outreach program. The peer educators also collected recommendations from the workers of their friends who might be interested in participating. All phone numbers thus collected were enrolled as participants in the quiz. There were a total of 2,494 participants enrolled at the start of the *FactoryQuiz* (360 in *Factory1*, 1,294 in *Factory2* and 840 in *Factory3*). Additional participants were enrolled after the start of the quiz (as discussed below).

5.2 Quiz Participants

The majority of quiz participants are believed to have only a primary school education. English is the national language and is taught in primary school, though local languages such as Ganda and Luganda are also commonly spoken in the areas where the quizzes were deployed. The quizzes were delivered only in English.

The literacy rate, defined as the proportion of the population over the age of 15 that can read and write English, is reported to be 66.8% in Uganda. Literacy is considerably higher for males (76.8%) than for females (57.7%) [25]. These figures may overestimate adult literacy rates since there is some evidence that literacy skills that go unexercised may be lost over time [26].

The majority of participants would be expected to have little or no experience with computers. The typical mobile phone used by participants in these two deployments had a color screen and SMS capability but no camera or data capability.

5.3 Procedure

The structure of the two deployments was as follows. The pool of participants was identified as described above. Then, an initial SMS message was sent to all the potential participants explaining the quiz and giving opt-out instructions. The TTC system sent a series of messages to those who did not opt out that included quiz content (i.e., questions about HIV/AIDS), a series of demographic questions (e.g., what is your gender?), messages to encourage continued participation, and announcements of the awarding of prizes. Participants were reminded via SMS about the possibility of winning airtime during the quiz; these reminders are believed to play an important role in maintaining participant motivation. In addition, TTC also sent informational messages, such as the location of local clinics that provide free HIV testing.

Messages were broadcast to all participants via SMS. The HIV/AIDS questions were sent at regular intervals over the course of two to three weeks, and responses were accepted any point between the initial distribution of a question and the completion of the quiz.

Multiple responses were allowed for all questions. TTC made this procedural decision in part because of the practice of phone sharing mentioned earlier. Additional support for this decision derives from their goal of promoting learning about HIV/AIDS: answering multiple times might be argued to provide additional opportunities for mastering question content. Feedback was given to all responses to content questions. If participants answered a question incorrectly, they received the correct answer embedded in a sentence. If they answered correctly, they were congratulated and also sent the correct answer in sentence form. At the end of each week, participants with the most correct answers were entered into a pool to win free mobile phone minutes. Additional prizes, such as football jerseys and mobile phones were awarded at the end. A message was sent at the end of the quiz to announce the completion of the quiz and thank the participants.

To participate in a quiz, one needs a mobile phone capable of receiving SMS messages on a specific broadly available carrier. Questions are sent from a single phone number (short code), and participants send a message back to the same number to reply. Messages sent and received from the quiz phone number are free for the duration of the quiz.

5.4 Interaction Details

The format of questions differed between the two deployments, but all questions had the same format within a deployment. The question format in the *DistrictQuiz* deployment was multiple-choice, with two to four alternatives for each question. Participants would be expected to achieve a score of 35% correct if they simply guessed at the answers. An example message, within the 160-character length limit, is as follows:

```
TTC Quiz Question: What causes AIDS? (1) A virus
called HIV (2) Witchcraft (3) Mosquitoes/insects.
Reply with "cause" and the number of your answer
```

The last part of the message, beginning with the word "reply," provides instructions on how participants must format their reply. When the response is received, the application back-end evaluates the response and immediately sends feedback. The system's response to an incorrect answer takes the form:

```
Too bad! The right answer is 1: AIDS is not caused
by witchcraft or insects; it is caused by the
"Human Immunodeficiency Virus" (HIV)
```

In the *FactoryQuiz* deployment the question format was true/false, thus participants would be expected to achieve a score of 50% correct by simply guessing. An example message is as follows:

```
TTC Quiz Question: HIV can be prevented by using
condoms correctly and consistently (1) true (2)
false. Reply with "condom" and the number of your
answer
```

[2] Note that because of the widely reported sharing of mobile phones in Africa [24], we cannot be sure that the invitation to participate reached the person whose use of the phone was originally detected.

The three factories received the same questions but in a different order and thus order and content effects can be disambiguated through analysis [27].

6. RESULTS

Each response had to begin with one of the quiz keywords in order to be identified as a response to the quiz. All responses from participants were entered into a database with one record per response. These records are comprised of the following information:

1. *Deployment name* – quiz deployment identifier
2. *User ID* – anonymized participant identifier
3. *Question number* – ordinal position of the question in the sequence of messages comprising the quiz
4. *Message* – participant SMS response
5. *Correct?* – automatic evaluation of participant message
6. *Time* – date and time of participant message

We then pre-processed the data to eliminate three types of responses. First, any record that contained a blank in the *message* field was judged to have been an incomplete response and thus eliminated (less than 1% of total responses). Second, the use of two particular keywords – “give” and “test” – in the quiz resulted in a number of apparently spurious text messages in the dataset. An example is “Give 9500 <phone number>”. This is the format used for transferring money to another person through a mobile money service [28]. We eliminated these responses, which amounted to a total of 2.4% of responses being excluded from the *Factory2* sample and 1.8% being excluded from the *DistrictQuiz* sample. (The other quizzes did not use either of these keywords.) Third, many participants (ranging from 16% to 24% in the four samples) responded more than once to a question. We only counted a respondent’s first response to a question in order to assess their starting knowledge.

We focus our analysis on two aspects of the data: accuracy and participation rates. Accuracy assesses participants’ knowledge of the content tested through the quizzes; this information is of primary interest to the TTC organization so that it can further tailor its interventions. We further analyzed incorrect responses to explore the effectiveness of SMS as UI. Participation rates across the two deployments are of interest as a way to begin to gain insight into how social factors might be leveraged to encourage participation in the quizzes *in lieu* of monetary incentives.

6.1 Accuracy of First Responses

Correctness was ascertained automatically by a scoring mechanism in the TTC system. The scoring algorithm operated as follows. Consider the question:

```
TTC Quiz Question: A healthy looking person can be
HIV positive. (1) true (2) false. Reply with
"healthy" and the number of your answer
```

In this instance, the system counted as correct any answer that took the form “healthy 1”, irrespective of any additional text that followed. More generally, for a response to be marked correct it must (a) follow the proper format (<keyword> <number>) and (b) <number> must correspond to the correct answer to the given question.

The results for the four samples are shown in Figure 2 and indicated by the lines labeled “base accuracy.” In all cases, accuracy was above chance levels. Accuracy is lowest for the *DistrictQuiz* sample; it begins at 40% for the first question, drops for question 5, and levels out at approximately 60% for the rest of the quiz. Base accuracy is considerably higher for the three factory samples that comprised the *FactoryQuiz* deployment as can be seen in the next three panels of Figure 3. Base accuracy for *Factories 1*, *2*, and *3* begins at roughly 60% and hovers around approximately 80% after the first few questions.

The base accuracy difference between the two deployments may be attributable to a number of factors. Recall that questions used a multiple-choice format in the *DistrictQuiz* (and correctness by chance was equal to 35%), while the *FactoryQuiz* deployment used a true/false format (where change is equal to 50%). Also, since the participants in each of the three deployments of the *FactoryQuiz* were collocated they might have had opportunities to discuss quiz questions – potentially bringing shared knowledge to bear. Question 5 in the *DistrictQuiz* proved to be particularly difficult: it asked about the incubation period required before a newly acquired HIV infection would be detected reliably (correct response is 30 days).

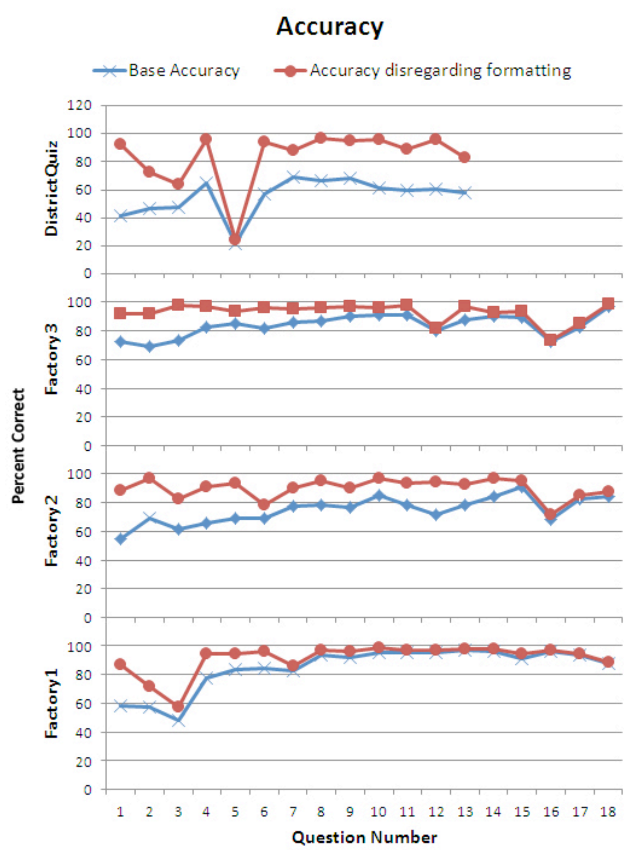

Figure 2: Percent correct for first responses to each question for *DistrictQuiz* and *FactoryQuiz* participants.

6.2 The Cost of Formatting Requirements

In order to better understand the challenges of SMS as UI, we looked at the difficulties participants had reproducing the required response format. To do this, we analyzed responses marked as incorrect by the TTC system and identified those responses that had correct content but did not follow the required format.

Specifically, for each message marked INCORRECT by the TTC System:

- If the correct answer is TRUE, we search the user's message for the following strings: "1", "true", "yes". If any of these strings are found, we assume the user's answer is correct but improperly formatted. Otherwise, we assume the answer is incorrect.
- If the correct answer is FALSE, we search the user's message for the following strings: "2", "false", "no". If any of these strings are found, we assume the user's answer is correct but improperly formatted. Otherwise, we assume the answer is incorrect.

The re-computed accuracy rates are shown in Figure 2, indicated by the lines labeled "accuracy disregarding formatting."

In all four cases, accuracy is considerably higher if the requirement of the precise formatting is disregarded. We can quantify this cost to accuracy incurred by the participant by calculating the average difference between the base accuracy rates and the rates when formatting is disregarded. These results are shown column 3 of Table 1. The cost ranges from slightly less than six percentage points to almost 31 percentage points. Participants in the *DistrictQuiz* sample showed a larger cost (30.92%) than participants in the three factory samples (5.97% to 14.10%). The *DistrictQuiz* cost was also more sustained than in the Factory quizzes, which can be seen in Figure 2. Specifically, the accuracy rates for the three *FactoryQuiz* samples converge within the first six questions, indicating that participants learned the format within that period. However, the two lines never converge in the case of the *DistrictQuiz*; the cost of requiring a specific format is still over 20% at the end of the quiz.

Table 1: Accuracy cost of requiring specific formatting in assessing correctness of answers.

Sample	Accuracy Disregarding Formatting	Base Accuracy	Accuracy Cost of Formatting Requirement
District	84.64%	53.72%	30.92%
Fact. 1	91.88%	85.91%	5.97%
Fact. 2	87.64%	73.53%	14.10%
Fact. 3	91.42%	82.89%	8.53%

We cannot definitively say why the *DistrictQuiz* participants had more difficulty than the FactoryQuiz participants in mastering the required formatting. Averaged over all participants in a sample, the DistrictQuiz participants made slightly more responses (median = 26) than any of the *FactoryQuiz* participants (median ranges from 19.5 to 20.5). This is not consistent with increased time on task as the explanation for the differences in formatting accuracy. As mentioned previously, it is possible that social learning among the presumably more connected participants in the factory deployments played a role. We note that the difficulty participants had mastering the required format is not surprising given previous research [16] which found that even with training, healthcare workers made errors when generating specifically formatted responses.

6.3 Causes of Formatting Errors

The cost of incorrect formatting to accuracy was substantial for all four samples, but the data shown in Figure 2 underestimate the degree to which participants failed to reproduce the modeled format because that analysis is based only on responses that contained the correct answer. Table 2 is based on the percentage of incorrectly formatted responses for the four samples irrespective of correctness of the content.

To better understand the nature of formatting challenges, we further analyzed the responses deemed incorrect by the TTC system to place them in one or both of the following categories: (a) "mechanical" format error or (b) response of a "conversational" nature.

- A *mechanical format error* is an answer that did not follow the formatting convention. If the keyword (e.g., "condom") and a valid answer (e.g., "2") appear anywhere in the message, it is considered to contain a mechanical error.
- A *conversational response* is one that contains more than four "words." (We define a word as a string of one or more characters separated by spaces.)

An example of what we consider a conversational response follows:

```
TTC Quiz Question: Family planning is an equal responsibility of both the wife and husband in the family. (1) true (2) false. Reply with "family" and the number of your answer
```

```
Observed reply: Family planning is the responsibility of both husband and wife
```

In this example, the participant produced the correct response but did not format it correctly.

Table 2: Conversational and mechanical errors for *DistrictQuiz* and *FactoryQuiz* participants. First Q/Last Q refers to the error rate on the first and last questions of the quiz; average is the mean error rate over the entire quiz.

Sample	Conversational Responses		Mechanical Formatting Errors	
	Average	First Q Last Q	Average	First Q Last Q
DistQuiz	9.70%	17.69% 8.55%	36.75%	57.31% 32.41%
Factory1	2.67%	5.70% 2.54%	7.80%	36.08% 1.75%
Factory2	3.74%	5.78% 1.80%	18.26%	43.68% 4.88%
Factory3	3.16%	4.49% 2.33%	10.28%	27.33% 2.5%

Table 2 reveals more detail about the pattern of errors. Participants in the four quizzes made almost four times as many mechanical errors as conversational ones. They also improved with more experience in the quiz (as shown by the first question / last question error rate comparison in columns 3 and 5). Although *DistrictQuiz* participants improved over the course of the quiz, by the end their responses were incorrectly formatted almost a third of the time (32.41%). This is consistent with the results shown in Figure 2.

DistrictQuiz participants produced the largest percentage of conversational responses (more than double that of the highest other sample). However, the relatively low rate of conversational errors overall was somewhat surprising because we expected the presumed lack of computer experience among the participants would result in more such responses across the board. The preponderance of mechanical formatting errors may indicate that participants understood responses needed to be in a particular format, but had trouble executing that format precisely.

A UNICEF program manager (one of TTC's partner organizations) with extensive experience deploying SMS based solutions to rural populations in developing countries explains the challenge this way: "It's hard for the average user to grasp the idea of interacting with a computer via SMS. Instead, folks assume they are interacting with another human – sometimes a very dumb human. Making the jump to the idea of a computer being on the other end is too far out of their everyday experience." [personal communication] Perhaps this explains a number of the mechanical and conversational errors we saw.

6.4 Participation Rates

We now examine whether there was any difference in participation rate between the deployments. *DistrictQuiz* offered various material incentives for participation. Those with the highest levels of participation over the duration of the quiz were entered into drawings for airtime, as well as football jerseys (expected to be quite a draw since football [American "soccer"] is very popular on the African continent). *FactoryQuiz* offered the same types of prizes and, in addition, "demand characteristics" [29] might have been operational as a result of management sanction of the quiz deployment. Because of this, workers might have been expected to encounter "peer pressure" to participate through their interactions with other workers who were also aware of the quiz deployment.

Participation rate is the number of individuals who answered a question relative to the number who had the opportunity to participate. We calculated the *base rates* of participation on a per question basis relative to the number of outgoing messages. In the case of *DistrictQuiz*, the denominator was the same throughout the experiment; all questions were sent to the entire sample identified at the start of the study (no one opted out of the study). In the case of *FactoryQuiz*, the number of outgoing messages increased over the course of the quiz as new participants were added. (New participants joined the quiz by sending an SMS to the short code for the quiz, which they could obtain from participating colleagues.) Results for the four samples are shown in Figure 3 and indicated by the lines labeled "base rate". Participation rates varied from a low of between 5% and 10% in *DistrictQuiz* to roughly 50% in of *Factory1* in *FactoryQuiz*.

The low participation rates for the *DistrictQuiz* sample may have resulted from a lack of context. While announcements were made via regional media outlets, participants may not have seen them and, therefore, not had sufficient understanding of the quiz (and/or TTC) to feel comfortable responding (particularly given the stigma associated with HIV/AIDS in parts of Uganda). In contrast, *FactoryQuiz* participants did have a context for expecting the quiz since it was socialized extensively in their place of employment. Consistent with this explanation is the finding reported above that some participants in *DistrictQuiz* indicated (via conversational response) that they wished the messages to stop, though they did not follow the opt-out instructions sent at the start of the quiz.

Since we were concerned that the opt-out instructions had not been effective and might have disproportionately penalized the *DistrictQuiz* sample, we computed a *conditional participation rate* for both quizzes. The conditional participation rate measures participation relative to the number of people who *implicitly* opt in to the study by responding to at least one question. The first week's conditional participation rate is 100%. Subsequent weeks are computed as follows: The denominator for the conditional participation rate computation for a given question in the quiz sequence is the running total of people who had participated in the quiz up to that point (not the total number of outgoing messages). To put it another way: Whenever a new person answers a question, the denominator is incremented by one.

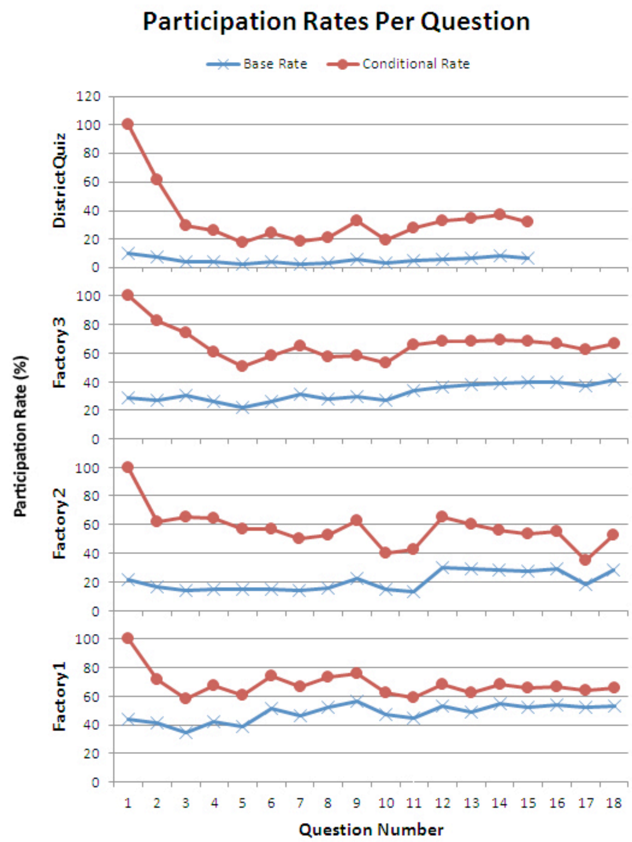

Figure 3: Participation rates for *DistrictQuiz* (Quiz 1), and the three factories in *FactoryQuiz* calculated as a base rate (top) or a conditional rate (bottom).

The conditional participation rates are indicated by the lines labeled "conditional rate" in Figure 3. By this method, the participation rate in all deployments is higher than the base rate (when the denominator equals the total number of messages sent by the TTC system), but the rate for the *DistrictQuiz* is still lower than that for the *FactoryQuiz* deployments. About one quarter of participants were retained after opting in to the *DistrictQuiz*, while a much larger percentage remain active in the samples from the *FactoryQuiz* deployments. Conditional participation by *Factories 1* and *3* is comparatively high (between 60% and 80%). Thus, if we restrict the sample to those who indicated interest in participating, the participation rates are considerably higher.

Higher participation rates are again observed for *FactoryQuiz* (as compared to *DistrictQuiz*) when we examine the number of questions each participant answered. More than 40% of the DistrictQuiz participants answered only one question, whereas the comparable rate for the three FactoryQuiz samples was 25% or less. More than 50% of the *DistrictQuiz* sample answered only one or two questions, whereas the comparable numbers for the three *FactoryQuiz* samples ranged from 5 to 9. These results are shown in Figure 4.

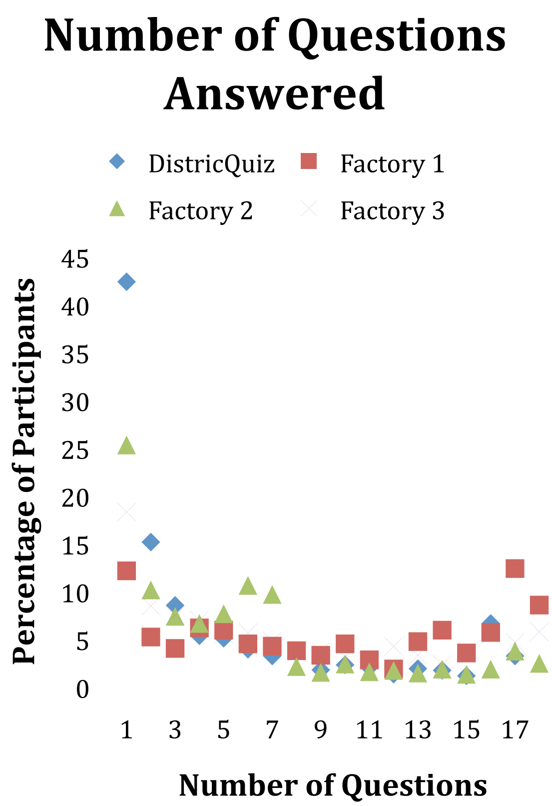

Figure 4: Number of questions answered by participants in the *DistrictQuiz* and the three factories in *FactoryQuiz*.

"Social recruitment" may be an effective way to increase participation rates. We noted above that new members joined the quiz in the *FactoryQuiz* in response to recruitment messages such as the following sent to already enrolled people:

```
Dear TTC quiz participant, remember that sending
an SMS is FREE with lots of prizes to win. Tell
everyone to join in by sending FACTORY2 YES to
8181
```

These messages were followed by a significant number of additional participants opting in to the quiz. Specifically, 41%, 67%, and 31% of the total opt-ins occurred in the two days immediately following the broadcast of such messages to workers in *Factories 1*, *2*, and *3* respectively. In total, 1846 participants joined in this manner.

7. DISCUSSION

Our analysis indicates participants in the two deployments know much of the content the quiz tested. This is an encouraging result with respect to the continuing spread of HIV/AIDS in Uganda. For instance, when formatting is disregarded, accuracy levels averaged over the entire quiz ranged from 84.6% to 91.9%. TTC also reported progress on meeting some of their broader goals. Specifically, health centers located in the *FactoryQuiz* towns experienced a three-fold increase in HIV test requests from workers. At least some of this increase likely resulted from the awareness brought about by the quiz deployments.

The analyses in this paper helped us make progress on the two research goals we stated at the beginning of this report: to explore the viability of SMS as a user interface technique for users who do not receive explicit training in the interface and to explore the potential impact of social factors on participation rates.

With respect to SMS as UI, we found that the formatting requirement does hamper respondents at first, but considerable learning took place over the course of the quiz. The rates of incorrectly formatted responses fell by approximately 90% for the three *FactoryQuiz* samples. The improvement was smaller (though still substantial) in the case of the *DistrictQuiz* sample, where the error decreased by more than 50%. The decreases in formatting errors that took place from the first to the last question in a quiz suggest that this learning may be mediated by additional experience, either reading more questions and/or formulating more responses. The greater learning in the three *FactoryQuiz* samples is consistent with a social learning explanation: greater presumed contact among quiz participants in the factory samples. There were not many conversational responses (just under 10% for the sample with the highest rate), suggesting that participants were able to adapt to the response format required by the TTC system's SMS user interface[3].

Thus, we conclude that structured SMS messages can be used effectively with untrained users in an application where errors are tolerable (not the case in data-critical applications as discussed in [16], [17] and [18]). Based on our analysis, though, we believe that it may be possible to help participants more quickly learn how to respond correctly to the TTC system by modifying the feedback it sends. The feedback could be based on the spreadsheet functions we developed for our analysis of responses originally marked as incorrect. Such functions could be incorporated into the TTC system to distinguish between content and formatting errors. This augmented system could then either send more specific error messages to participants in order to help them conform to the prescribed message format. Alternately, the new system could directly ascertain the correctness of messages that do not completely comply with the prescribed form.

Our second research goal was to investigate the effect of social structures on participation rates. Here we made progress, though the results must remain provisional. While this data does not allow us to unequivocally attribute the difference between the results in the *DistrictQuiz* and those in the *FactoryQuiz* to social factors (either demand characteristics exerted by embedding the quiz in the participants' place of employment and/or due to discussions with or pressure from peers), we are encouraged by the results and will work with TTC in future quiz deployments in a way that allows us to be more clearly test the potential impact of social factors on scaling participation.

The question about the impact of literacy remains unanswered. Those who responded were largely able to make effective use of structured SMS and only a small number of responses could not be deciphered (a subset of the ones marked as having no response). What we do not know is if some of the approximately 75% who did not respond to the quiz in the *DistrictQuiz* sample

[3] The requirement that responses begin with a keyword in order to be recognized as a response may have resulted in an undercounting of this category of response.

may have lacked the necessary literacy for the quiz. All we can say for sure is that overall those who did participate had sufficient literacy for the task.

One final note about one aspect of the protocol adopted in the quizzes, namely requiring targeted participants to opt out of the quiz in order to cease receiving quiz questions. We found conversational responses in the data that asked for the quiz questions to stop, however none of those users used the prescribed message to officially opt out. Thus, we believe this approach was not successful. In addition, our conditional participation analysis supports the thesis that an opt-in strategy is likely to improve participation rates. While opt-in vs. opt-out is an active research topic in the context of online privacy policies [30, 31], we have been unable to find previous research that is directly applicable to our situation. However, we believe that the potential for annoyance due to repeatedly sending unsolicited messages clearly argues for using an opt-in rather than an opt-out strategy in our case.

8. FUTURE WORK

As this partnership between TTC in Africa and researchers in the USA moves forward, we will explore ways social computing capabilities can make TTC's quizzes more scalable and sustainable. We plan to carry out a larger study at seven factories in rural Uganda where we will deploy social incentives in a controlled manner. Specifically, we will deploy three conditions (individual-focused, group-focused, and combined incentives) in order to study the effectiveness of different kinds of social incentives. We also hope to incorporate changes to the system design based on the findings reported here.

9. ACKNOWLEDGMENTS

Many thanks to the participants in the TTC quizzes and to TTC's partners in the two deployments.

10. REFERENCES

[1] WHO World Health Statistics 2010. http://www.who.int/whosis/whostat/2010/en/index.html

[2] UN AIDS Data 2008. http://www.unaids.org/en/KnowledgeCentre/HIVData/

[3] Wikipedia: Community Health Worker. http://en.wikipedia.org/wiki/Community_health_worker

[4] AMREF: Community health workers the answer to Africa's problems. http://uk.amref.org/news/community-health-workers-the-answer-to-africas-problems-/

[5] Blycroft Limited. African Mobile Factbook 2008.

[6] Jensen, R. The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector. The Quarterly Journal of Economics, 3 (2007), 879-924.

[7] Corbett, S. Can the Cellphone Help End Global Poverty? The New York Times (2008).

[8] Donner, J. (2007). The Rules of Beeping: Exchanging Messages Via Intentional "Missed Calls" on Mobile Phones. Journal of Computer-Mediated Communication, 13(1), article 1, http://jcmc.indiana.edu/vol13/issue1/donner.html

[9] Vital Wave Consulting. mHeatlh for Development: The Opportunity of Mobile Technology for Healthcare in the Developing World. UN Foundation - Vodafone Foundation Partnership, Washington, DC and Berkshire, UK, 2009.

[10] EpiHandy. http://www.epihandy.org/index.php/Main_Page

[11] OpenXData. http://www.openxdata.org/

[12] Staff Reporter. MXit registrations surpass 14 million. Broadband News South Africa (2009).

[13] Kechice, B. Leveraging SMS to Attract the Low End of the Income Scale. Africa & Middle East Telecom Insider (2009).

[14] FrontlineSMS. http://www.frontlinesms.com/

[15] RapidSMS. http://www.rapidsms.org/

[16] Berg, M., Wariero, J. & Modi, V. Every Child Counts - the Use of SMS in Kenya to Support the Community Based Management of Acute Malnutrition and Malaria in Children Under Five. Report downloaded at: http://www.childcount.org/reports/ChildCount_Kenya_InitialReport.pdf

[17] Eagle, N. Africa and the Mobile Phone. http://www.thesolutionsjournal.com/node/623

[18] ELMR: Lightweight Mobile Health Records. Kumar, A., Purandare, A., Chen, J., Meacham, A., Subramanian, L. SOGMOD '09, Providence, RI, June 29-July 2, 2009.

[19] Ushahidi. http://www.ushahidi.com/

[20] Project EPIC: Tweak the Tweet. http://epic.cs.colorado.edu/helping_haiti_tweak_the_twe.html

[21] HealthLine. http://www.cs.cmu.edu/~healthline/

[22] Agarwal, S.K, Jain, A., Kumar, A., Nanavati, A.A., Rajput,N. (2010). The spoken web: A web for the underpriviledged. ACM Sigweb Newsletter, summer.

[23] Plauche, M. & Nallasamy, U. Speech Interfaces for Equitable Access to Information Technlogy. Information Technologies and International Development, Volume 4, Issue 1,2007, pp 69-86.

[24] Chipchase, J. Shared Phone Use (2006), http://www.janchipchase.com/sharedphoneuse

[25] CIA World Factbook – Literacy. https://www.cia.gov/library/publications/the-world-factbook/fields/2103.html

[26] Willms, D. J. & Murray, S. Gaining and Losing Literacy Skills Over the Lifecourse. Statistics Canada, 2003.

[27] McBurney, D.H. & White, T.L. Research Methods, 8th edition, 2009, p. 278.

[28] Leishman, P. Understanding the Mobile Money Customer: an Interview with John the Taxi Driver, http://mmublog.org/africa-east/understanding-the-mobile-money-customer-an-interview-with-john-the-taxi-driver/

[29] Psychology Glossary: Demand characteristics. http://www.psychology-lexicon.com/cms/glossary/glossary-d/demand-characteristics.html

[30] Lai, Y-L &Hui, K-L. Internet opt-in and opt-out: Investigating the roles of frames, defaults and privacy concerns. SIGMIS CPR '06.

[31] Bellman, E., Johnson, E.J., & Lohse, G. L. (2001). Onsite: to opt-in or opt-out?: It depends on the question. Comm. of the ACM, 44(2).