Voice-Based Quizzes for Measuring Knowledge Retention in Under-Connected Populations

Agha Ali Raza
ITU, Lahore, Pakistan
agha.ali.raza@itu.edu.pk

Zain Tariq*
ITU, Lahore, Pakistan
zain.tariq@itu.edu.pk

Shan Randhawa
ITU, Lahore, Pakistan
shan.rand@itu.edu.pk

Bilal Saleem
ITU, Lahore, Pakistan
bilal.saleem@itu.edu.pk

Awais Athar
EMBL-EBI, Cambridge, UK
awais@ebi.ac.uk

Umar Saif
ITU, Lahore, Pakistan
umar@itu.edu.pk

Roni Rosenfeld
CMU, Pittsburgh, USA
roni@cs.cmu.edu

ABSTRACT

Information dissemination using automated phone calls allows reaching low-literate and tech-naïve populations. Open challenges include rapid verification of expected knowledge gaps in the community, dissemination of specific information to address these gaps, and follow-up measurement of knowledge retention. We report *Sawaal*, a voice-based telephone service that uses audio-quizzes to address these challenges. Sawaal allows its open community of users to post and attempt multiple-choice questions and to vote and comment on them. Sawaal spreads virally as users challenge friends to quiz competitions. Administrator-posted questions allow confirming specific knowledge gaps, spreading correct information and measuring knowledge retention via rephrased, repeated questions. In 14 weeks and with no advertisement, Sawaal reached 3,433 users (120,119 calls) in Pakistan, who contributed 13,276 questions that were attempted 455,158 times by 2,027 users. Knowledge retention remained significant for up to two weeks. Surveys revealed that 71% of the mostly low-literate, young, male users were blind.

CCS CONCEPTS

• **Human-centered computing → Sound-based input / output**.

KEYWORDS

HCI4D; IVR; low-literate; blind; mobile phone; impact.

*The first two authors contributed equally to this paper.

CHI 2019, May 4–9, 2019, Glasgow, Scotland Uk

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300642

ACM Reference Format:
Agha Ali Raza, Zain Tariq, Shan Randhawa, Bilal Saleem, Awais Athar, Umar Saif, and Roni Rosenfeld. 2019. Voice-Based Quizzes for Measuring Knowledge Retention in Under-Connected Populations. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3290605.3300642

1 INTRODUCTION

Voice-based services over simple mobile phones have emerged as a viable way to rapidly disseminate useful and time-sensitive information to under-connected populations (low-literate, poor, linguistic minorities, blind, women). Such services are being used in developing regions around the globe for e-government [53], Citizen Journalism [4, 21, 39], disaster response and recovery [52], prevention, monitoring and control of epidemics [67] and for providing access to information regarding health [56, 67], finance [53], agriculture [42, 43, 60], and job search [50, 65].

Generally, impact assessment of information campaigns is carried out via follow up surveys to measure knowledge retention [10, 11, 40, 66]. However, impact assessment presents a unique set of challenges when information is mass disseminated using speech over simple phones. To encourage spread, inclusiveness, and anonymity, such services do not require users to go through any formal recruitment or registration and a significant fraction of them only engage with a handful of calls. As a result, manual or telephonic baseline and end-line surveys are not feasible. Even where feasible, such surveys are hard to scale, prone to delays and resource hungry. Consequently, despite a growing number of telephonic campaigns, there is little focus on the measurement of retention of the delivered information. Due to a lack of rapid, scalable and reliable mechanisms to quantify and measure knowledge retention, campaigns mostly resort to measuring only the extent of delivered information (for instance, number of calls and number of users who listened to the information content). Assessment of knowledge-impact is only a piece of the bigger puzzle of achieving behavior change, which is still harder to quantify and measure.

Taking a step back, it is difficult even to rapidly measure the baseline knowledge of an under-connected community about specific topics. The first step towards planning an information campaign is to find out the current levels of awareness, gaps in knowledge and prevalent misconceptions in the target community that may guide the development of the information content and delivery mechanisms. It is desirable to find out any taboos and cultural hurdles that may hinder the objectives of the campaign. Currently, there is no quick and easy way of doing this.

Our approach is to merge knowledge-gap discovery (about specific, known or hypothesized topics), information dissemination and knowledge retention measurement into one service, using voice-based quizzes. Our **research question** is: How can a platform be created that reaches and engages under-connected populations, in order to:

(1) discover existing gaps in their knowledge regarding various known topics-of-interest, rapidly and at scale,
(2) disseminate relevant and specific information to fill their knowledge gaps, and
(3) motivate them to keep interacting with the platform for several days to allow measuring retention of the delivered information.

We show that voice-based quizzes over simple mobile phones, consisting of multiple-choice questions, can be used to simultaneously measure the existing knowledge gaps (around specific topics) as well as to disseminate information. The same questions can be rephrased and repeated at regular intervals to measure retention of conveyed information. Long-term user engagement can be encouraged by allowing users to contribute questions, and through the use of social connectivity, gamification and spirit of competition to make the service engaging and fun for the target audience.

Our quizzes spanned topics including general knowledge, physical and mental health, hygiene, child safety, water and sanitation, people with special needs, family planning, environment, constitutional rights, and law. We made the design choice to include diverse information genres as we did not want our initial findings to be specific to a particular genre and wanted them to apply more generally. Common themes among user-contributed questions included history, literature, religion, sports, general knowledge, geography, and current affairs. In the next section, we review related literature and situate our contribution accordingly.

2 RELATED WORK

This section reviews voice-based services (also called Interactive Voice Response (IVR) services) for under-connected populations and use of quizzes and games for knowledge measurement over other modalities. We did not find any voice-based telephone services for large-scale knowledge-retention measurement among under-connected populations.

IVR Services for Under-connected Populations

IVR services are being utilized for providing useful services to under-connected populations. Such services have been used to leverage social connectivity for creating social networks, crowd-sourcing and data collection platforms, and information dissemination platforms. In India, Avaaj Otalo [43] provided farmers a speech platform to discuss their issues with other members of the agricultural community. Sangeet Swara [61], a voice-based community forum for low-literate communities in India, allowed sharing songs, jokes, and poetry, and utilized community moderation to filter and rank the content. Baang [49], another voice-based social platform, was deployed in Pakistan that allowed under-connected populations to share audio content of diverse genres. User voting, comments, and sharing were shown to lead to organic spread, more engagement, higher retention and true dialog among participants. Such platforms could substitute online social media for low-income, low-literate, visually-impaired communities [62]. IVR platforms are also being used for large-scale, rapid speech data collection [29, 46]. Voice-based micro-tasks have been used for crowd-sourcing in low-income [63] and visually impaired [64] settings. Gurgaon idol [26] is a singing competition in collaboration with a community radio station in which community members call an Interactive Voice Response (IVR) service to record songs, and vote to select the best songs. VideoKheti [18] uses a multi-modal approach to deliver useful content to its users and suggests that although speech interfaces do not always work for low-literates, they are still more useful than graphical and textual interfaces. Healthline [56] was deployed in rural Sindh, Pakistan to aid the efforts of community health workers by providing them with a speech-enabled information service. Raza et al. designed Polly, an IVR service that utilized entertainment to rapidly gain traction among low-literates in Pakistan [47] and India [48], and was also used to deliver employment-related information [51]. CGNet Swara [39], a citizen journalism service in India allows users to share news and problems. Gramvaani [4] has been using its IVR systems for social responsibility and awareness campaigns since 2009. A study based on users from different developing regions (India, Africa, Philippines) suggests that speech-based systems perform better than other interfaces (textual, graphical) for low-literate users [36].

For audio-visual modality, [37] presents a social networking application for low-literate farmers in India. Bidwell et al. report an audio repository accessible over smartphones for oral elders in rural Africa [14]. Asynchronous audio [38] and video [45] content has also been employed to assess questions and identify knowledge gaps in the health domain.

We build upon the literature by employing user-contributed content, gamification, and social connectivity to engage and

retain our audience [49, 61] and promote the viral spread of the service [47, 49]. Community-questions enable discovering community's topics of interest and allow users to repurpose the platform to suit their information needs (similar to the social component of Avaaj Otalo [43]). Such content can also potentially uncover prevalent misconceptions. The novelty of Sawaal is that it allows information dissemination as well as rapid measurement of its retention by engaging users with multiple-choice questions. This is of practical use for HCI researchers and practitioners who employ IVR services for large-scale information dissemination and want to measure knowledge-impact beyond a simple count of calls and users.

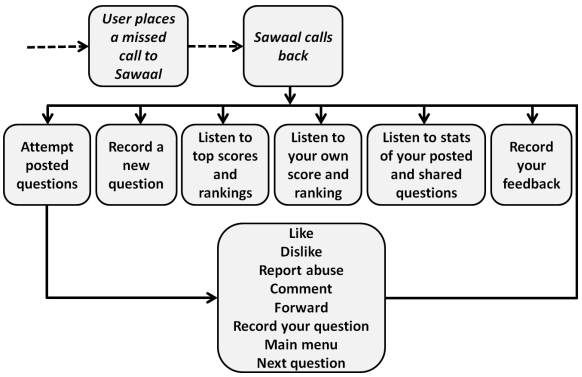

Figure 1: User Interface Work-flow

Learning and Knowledge Measurement

Active learning, when compared to passive, has been shown to lead to better learning outcomes, quicker transfer of knowledge, more engagement with the content and other participants, higher retention, and behavior change [12, 22, 31, 33, 57]. Active learning focusses more on learner participation, engagement with peers, and taking part in teaching the newly learned content, as compared to the unidirectional flow of knowledge in passive learning. Gamification and collaborative learning have been shown to promote active learning [15, 59, 68]. Gamification is being utilized in recent research on second language teaching and measurement of speaking skills of users [13, 19, 20, 27, 41]. Most of these games assume some level of literacy either in terms of education or computer, smartphone, web or Internet usage. Kam et al. use mobile video games to incentivize learning among rural children [24] and to improve English as a second language (ESL) literacy in India [25]. They present a receptive-practice-activation conceptual model [24] where the receptive phase teaches users about a particular topic, the practice phase provides more exposure to the subject being taught, and knowledge is tested in the activation phase. They also present the PACE (Pattern, Activity, Curriculum, Exercise) framework where users are repeatedly shown words with their meanings and are tested between the repetitions [25]. Learning is measured by comparing pre-test and post-test results. Multiplayer games [44] in competitive and collaborative modes are studied to test English language vocabulary retention in underprivileged school children in India. Mobile apps like Magoosh [5] and Duolingo [3] help smartphone users improve their vocabulary and language skills. Words are repeated periodically to gauge the memory (learning) of the users. Sporcle [6] provides internet users with a large set of quizzes to play and also create their own quizzes. The most notable use of IVR in training/learning is by CapacityPlus [1] in Kenya to train low-literate medical health professionals using a Q&A based IVR system. Participants take part in a lecture series and then test their knowledge using the IVR system. Brown et al. [16], measure the understanding of users about water contamination using a multi-modal mobile app. Sharing of animated videos over mobile phones to disseminate knowledge among low-literate farmers is studied in [32]. Farmers' impression of content difficulty and application of key-concepts are used to gauge learning outcomes. Medhi et al. [34, 35] study the correlation between low-literacy with learning ability using video content. They recommend sensitivity to cognitive differences beyond the inability to read when designing content for users with limited education.

We build upon the themes of repetition and active-collaborative style of learning [3, 6, 15, 24, 25, 34, 35] for better learning outcomes and measuring knowledge-retention. We extend the literature to a new modality: voice over simple phones. **The novel HCI contribution** of our work is that we bring together these themes to create a voice-based, community-driven question-answer platform for tech-novice users designed for measuring community's knowledge about specific topics, information dissemination and assessment of knowledge retention. Via administrator-posted questions, Sawaal allows measurement of community's knowledge about topics of interest (baseline), dissemination of information (via correct answers to the posted questions), and measurement of knowledge retention (via asking rephrased versions of the attempted questions after predefined intervals).

3 METHODOLOGY AND DESIGN

In this section, we describe the interface and design of Sawaal and our data analysis methods.

Interface Design

Sawaal (which means "Question" in Urdu) is a voice-based, telephone-based quiz service that allows users to attempt multiple-choice questions posted by others and to post questions of their own. Users can up-vote and down-vote the questions, post audio comments, compete against others for better scores and challenge their friends by forwarding interesting questions to them. Sawaal is accessible on simple

phones via voice calls and supports a simple interface to allow low-literate and non-tech savvy users to use it with ease. This makes it an inclusive information sharing platform.

In order to subsidize airtime cost for the users, Sawaal operates via missed calls: The interaction starts when a user places a call to Sawaal's phone number. As shown in Figure 1, Sawaal rejects the call (without answering) and plays a ring-back tone informing the caller to expect a call-back within few minutes. As Sawaal calls back, the interaction starts with culturally appropriate greetings, immediately followed by a randomly selected interesting fact. These facts are chosen to be fun, awe-inspiring and interesting and are intended to grab the user's interest early on during the interaction. This is followed by a brief introduction of Sawaal as a forum that allows users to post questions, answer posted questions and to challenge their friends to quiz competitions. Next, we play disclaimers highlighting the following points:

- Sawaal is a social platform and users are expected to use it politely and responsibly. Any profanity, hate speech or otherwise inappropriate content will be removed immediately and responsible users would be banned from the service.
- Sawaal takes no responsibility for the authenticity of user-posted information and users are advised to verify information before acting upon it.
- All content posted on Sawaal will be available to the public and will also be used for research purposes. Users must never post any personal information like phone numbers, addresses etc.

First-time callers are taken to a round of warm-up questions (10 questions on fun facts). The goal of this round is to help new users understand the concept of Sawaal and to learn the ropes of the user interface. To reduce cognitive load, help is provided to the users to navigate through the interface while making sure that the asked questions are funny and fascinating. This round does not allow users to vote or comment on the questions. After attempting all warm-up questions, users are asked to record their names, congratulated at the successful completion of their training, and then taken to the main menu of Sawaal. Return users do not get the training round and are taken directly to the main menu.

Sawaal provides a rich menu of options to users with simple choices towards the beginning (to facilitate new users) and sophisticated choices towards the end. The menu allows users: (1) to attempt questions posted by others, (2) to post new questions, (3) to listen to stats and rankings of top scorers, based on most correct answers and most up-voted questions, (4) to access their own profile, rank and score, (5) to listen to stats (likes, dislikes, reports, comments, number of correct/incorrect attempts) of questions previously posted attempted or forwarded to others by them, and (6) to provide feedback and suggestions.

Users who choose to **post a new question** are reminded of the ethics of posting and are advised not to share any personal information. They are allowed up to one minute to record a question, followed by 10 seconds to record each of four answer options. Finally, they are asked to mark one of the options as the correct answer. At this point, they can also forward this question to friends by entering their phone numbers. Sawaal calls up friends to deliver the challenge. It plays them the name of the sender, introduces Sawaal, and allows them to answer the shared question, after which they can browse Sawaal like a regular caller. Users are not allowed to answer their own posted questions. Although questions get published without moderation, moderators listen to content that is reported for abuse by users and remove it if found inappropriate

Users who choose to **attempt questions** posted by others are presented with a mix of administrator-posted and community-generated questions (algorithm explained in the next section). Questions cannot be skipped. Applause sound effects are played for a correct answer. For an incorrect answer, melancholic sound effects are played and the user is informed about the correct answer. After each correct answer, users are asked to mark the question as being easy, normal or hard. After each question, users are given options to (1) proceed to the next question, (2) listen to and post comments on the question, (3) up-vote, (3) down-vote or (4) report abuse, (5) forward it to friends (6) post their own question, or (7) go to the main menu. While attempting a question, users can hear its statement and answer choices repeatedly, but once answered, the same question never gets presented to them again.

We did not allow a skip option as it may nudge users into just answering all the fun questions and skipping most of the useful and important ones. We want users to think over and attempt all questions. We limited Sawaal to multiple choice questions, as opposed to open-ended ones, as (1) existing platforms (e.g. Baang [49], Sangeet Swara [61]) already allow this, and (2) we wanted to restrict the scope of the initial deployment as subjective, open-ended answers are harder to process automatically and require human moderation.

Sawaal keeps track of the scores of users (number of correct answers and attempted questions) and the popularity (likes) of the questions they have posted. It announces rankings of top 5 users with the most correct answers and top 5 users who have posted the most popular questions. Both of these rankings are available as daily, weekly and all-time favorites and can be accessed from the main menu. Users can also browse their own score and ranking as well as track stats of challenges that they throw to friends via forwarded questions. Finally, the main menu also allows users to provide their feedback and suggestions as unstructured audio recordings.

Order of Presentation of Questions

Sawaal has three types of questions: *Community-generated Questions* (CGQs), *Administrator-posted Questions* (APQs) and the initial 10 *Warm-up Questions* (WQs). The APQs are extracted by our team from authentic sources (as discussed in section *Administrator Posted Questions*) to gauge the knowledge of our audience regarding various topics of interest. Further, in order to measure knowledge-retention, each APQ also has a rephrased version. We will refer to these two variants of each APQ as *original questions* (OQs) and *rephrased questions* (RQs). We rephrase questions instead of repeating them verbatim to make sure we do not confuse true learning with memorization and rote learning. The rephrased questions are released over several days after the user has attempted the corresponding original questions. This allows us to track if there is any improvement/change in a user's knowledge about that question after they are informed of the correct answer. The *slow-release* allows us to *track temporal knowledge retention* i.e. for how long after attempting the original question do the users retain their knowledge of it? The order of presentation of questions is designed:

- To keep the users engaged and entertained and provide them with a mixture of CGQs and APQs.
- To make sure that users get exposed to several OQs quickly so that their respective RQs could be slowly released over several days.
- To cater to the needs of both steady users (who call occasionally and answer a handful of questions in each call) and enthusiastic users (who call frequently and attempt several questions in each call) and to make sure we have new questions when they call.

Algorithm: The order of presentation of questions attempts to cater to the aforementioned requirements. In the beginning, each user is presented with 10 warm-up questions. Users' responses at this stage do not count towards their score and a lot of guidance is provided. After the completion of training, Sawaal starts presenting users with a mixture of administrative and community questions. The questions are presented in sets of 8. As discussed, each administrator question has two rephrased versions: OQ and RQ. Initially, the algorithm releases more OQs mixed with a handful of community questions. This ensures that users get to hear all the OQs rather quickly. Sawaal also plays the correct answer each time users attempt a question incorrectly so that users know the correct answer once they have attempted a question. Each question is played to a user only once.

Once the required number of OQs have been attempted, the algorithm starts releasing RQs and we start getting the data required to measure knowledge retention of users. When users hang up and call back later, the sequence of questions continues from where it got interrupted. When the system

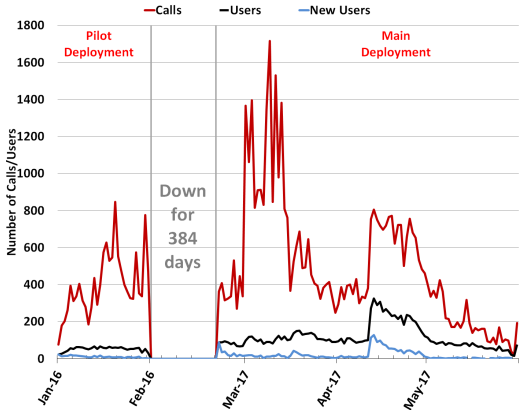

Figure 2: Calls, distinct users, and new users

runs out of OQs or RQs, it starts presenting users with just the CGQs. If it runs out of all questions, an apology prompt is played explaining that the system does not have more questions at the time and users should check back later.

The order of presentation of OQs is predefined but completely random as it is not based on any topic grouping, difficulty level or order of generation. This order remains the same for all users so that we can compare learning outcomes across them. To ensure that we are able to measure short-term and long-term knowledge retention, the rephrased questions are not released all at once but over several days. To achieve a uniform spread, an equal number of RQs are released immediately, after 24 hours, after 48 hours, and after 7 days of when their corresponding OQs are attempted.

For example, a user calls and attempts four OQs: OQ_1, OQ_2, OQ_3 and OQ_4. They will get to hear the corresponding RQ_1 immediately (if they stay on the same call or call back later), RQ_2 will not be released until 24 hours have elapsed from the time they attempted OQ_2, RQ_3 will not be released until 48 hours have elapsed and RQ_4 will be released after 7 days. Once released, the RQs never expire. This only guarantees that the RQs are attempted *at least* after the predefined intervals of the corresponding OQs.

The CGQs are sorted and presented using a mixture of popularity and recency. The most up-voted questions posted over the last 24 hours are presented first until we run out of them, next the most up-voted questions in the last 7 days are presented followed by all-time up-voted questions. A balance between popularity and recency and the fact that users hear a particular question only once ensures that users get high-quality questions to attempt but at the same time new questions do not suffer starvation.

Analysis Methodology

To answer the research question about reaching and engaging under-connected populations, we present usage statistics:

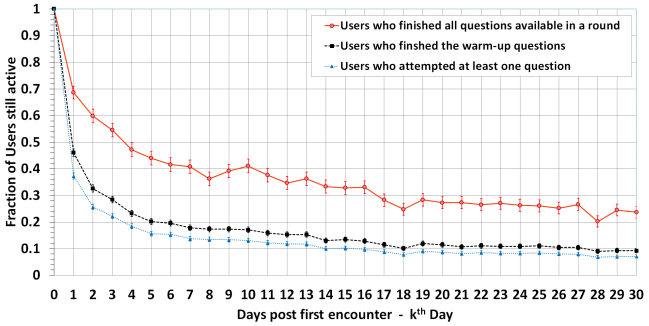

Figure 3: User Retention

extent of spread of the service, user retention, and use of various interface features. We also present user demographics from telephonic user surveys. Next, we look at the design of administrator-posted development-related questions and the various knowledge gaps that we found in the community using them. We measure and present knowledge retention both quantitatively and qualitatively. At this point, we also present some interesting observations. Next, we look at community generated questions and how they reveal users' information needs and preferences. Finally, to establish usability we present an analysis of user feedback and comments.

4 RESULTS AND ANALYSIS

Deployment

Sawaal was *never advertised or seeded*. We just made it available on a phone number that was in the same series as the numbers of two popular voice-based telephone services in Pakistan (The other services had their phone numbers ending in x111 and x115, while Sawaal was made available on x114). Almost immediately people discovered the new service and started calling. The first (pilot) version of Sawaal did not have rephrased questions. This version remained up for 31 days and received 8,909 calls from 578 users who contributed 3,176 questions that were attempted 73,467 times by 395 users. Questions were liked 45,013, disliked 21,108 times and rated as hard, medium or easy 38,572 times. We turned Sawaal off after this successful pilot for a year while we arranged for funding to support a longer deployment.

Sawaal was made live again on its old phone number, without any advertisement, for 101 days (3.4 months). During this period 3,433 users interacted with it via 120,119 calls, resulting in 13,276 community-generated questions (recorded by 807 users). Of these, 10,746 were valid (contained all four options) that were attempted 455,148 times by 2,027 users. The questions were answered correctly 211,242 times by 1803 users while incorrectly 243,905 times by 1,941 users. The questions were liked 6,999 times (by 489 users), disliked 1,805 times (by 303 users) and reported for abuse 24,643 times (by 586 users). 10,622 questions were rated as hard, medium or easy 203,673 times by 1,890 users. 667 users posted 6,296 voice comments on 3,291 questions, while 833 users shared (forwarded) questions 21,393 times with 2,115 friends. Figure 2 shows the number of calls, users and new users of the two deployments. Over the two deployments, users spent 5,130 hours interacting with Sawaal. On average, each user spent 2.23 hours on Sawaal, attempted 224 questions, contributed 7 questions, caste 45 votes (likes, dislikes, or reports), posted 3 comments and forwarded 11 questions to friends.

We also had 184 administrator-posted questions (92 OQs + 92 RQs). These were attempted 65,461 times (by 1,321 users), correctly answered 32,954 times (by 1,206 users) and incorrectly answered 32,507 times (by 1,278 users). 1,133 comments were posted on these questions by 499 users. They were liked 916 times, disliked 230 times, reported 706 times and forwarded 5,439 times by 281 users to 498 friends.

User Retention

It is necessary to engage a significant fraction of users for several days in order to meaningfully impart knowledge and measure knowledge retention. Figure 3 shows the user retention of Sawaal over time. We have considered three user groups for this analysis. Of our 3,433 over, 2,028 (59%) attempted at least one question. Of these, 1,406 users ended up completing all warm-up questions. And 381 users finished off all available questions in the system at least once. Therefore, the only users that we have not considered for this analysis are the ones who did not engage at all with the system and did not even attempt a single question (a downside of our lengthy disclaimer is that it saw a call drop-off rate of 28.6%).

Figure 3 shows the fraction of users who continue using Sawaal k days after their initial interaction (where k=0 to 30). The denominator only counts the subset of users who had a chance of using Sawaal on their k^{th} day. For example, a user who starts using Sawaal three days (72 hours) before the end of deployment only had a chance to use Sawaal for three days and consequently will not be considered in the denominator of the fraction of users who use the service 4 days or more after their initial interaction. Error bars marked on each graph show standard error.

We see high retention even among the casual users: 37% return to Sawaal on their second day, 25% on their third day, 14% after a week while 7% after a month. This number does not drop any further for several weeks and even after two months 6.9% of all users still keep calling (graph not shown). Therefore, we have access to the bulk of users in the first few days and a smaller yet significant fraction of users for several months. Larger fractions of users, who get engaged early on, end up returning to the service as shown by the other two graphs. After 30 days of initial use, we have access to 9% of all trained users and 23.7% of all motivated users.

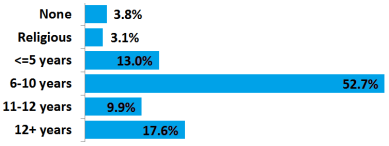

Figure 4: Formal Education: From 131 Survey Participants

Use of various Interface Features

Figure 5 shows the overlap among sets of users who engage with various interface features of Sawaal, using an *upset graph* [30]. As shown, 39% of all users just answer posted questions without engaging with any other activity. The second highest number (18%) engage with all activities. Interestingly this is followed by 10% users who attempt questions and also share questions with others.

Viral Uptake of the Service

As discussed in the section on *Deployment*, Sawaal was never advertised. It has only two mechanisms of spread: direct viral spread via forwarded questions and indirect viral spread via word-of-mouth where users inform others about Sawaal offline. We find that 51% of all users (N=3,433) had their first interaction with Sawaal via delivery calls (calls in which Sawaal delivers users' forwarded questions to their friends). Therefore the question/challenge sharing mechanism was responsible for bringing in half of all users. Remaining users found out about Sawaal via word-of-mouth.

User Demographics

To know more about our users, we conducted telephonic interviews of 131 randomly chosen users of Sawaal. Of these, 8% were identified as females. This is consistent with similar services e.g. Sangeet Swara (6% females) [61], Baang (10%) [49], Polly (11%) [51], CG Net Swara (12%) [39], and Ila Dhageyso (15%) [21]. 71.32% of the survey participants were visually impaired. This was an interesting finding as we had never advertised Sawaal and had never tried to publicize it to blind people. Users came from all provinces of Pakistan and 54 different cities and villages. However, most of them

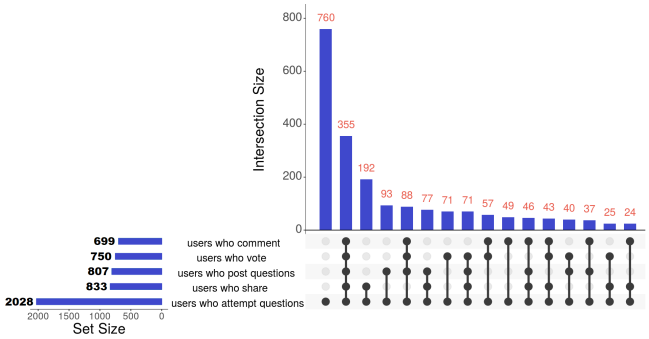

Figure 5: Use of various interface features

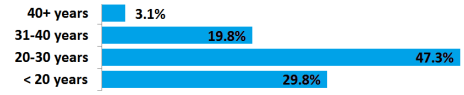

Figure 6: Age: From 131 Survey Participants

were from Punjab (64%), Khaiber-Pakhtunkhwa (18%) and Sindh (13%). Participants were very diverse in terms of preferred languages (Punjabi: 31.5%, Urdu: 22.3%, Pashto: 18.5%, Saraiki: 15.4%, Hindko: 4.6%, Balochi: 3.8%, Sindhi: 3.1%, and Potohari: 0.8%). Comparatively, there is much less language diversity in recorded questions (in 906 randomly-selected recordings: Urdu: 90%, Punjabi, Pashto, Saraiki: 0.2% each). Possible reasons for this bias include users trying to ensure that more people understand their recorded questions, questions in regional languages getting intelligibility complains from others in comments, and popular image of Sawaal as an Urdu service as evidenced by the interface language.

As shown in Figure 6, 30% of the users were less than 20 years old. These young users were mostly school pupils who were enthusiasts of Sawaal and claimed that it helps them prepare for their exams. Highest fraction (47%) of the users were of ages 21–30 and 52% had 6–10 years of education. At the same time, there were 7% users who had no formal education and another 13% who had 5 years of education or less. Also of note are 28% users with more than 12 years of education. By profession, 38% of users were students, 20% were unemployed, and 10% earned their living through manual labor. Others included teachers (6%), government employees (5%), small business owners (4%), telephone operators (4%), priests (3%) with a handful of farmers, housewives, drivers, gardeners, mechanics, and policemen.

Only 17% of the participants owned a smart phone while 35% reported having ever had used the Internet, Facebook or Whatsapp. 80% of users reported spending PKR 600 (USD 5.7) or less on monthly mobile airtime. 68% said that they would stop using the service if they have to pay for the airtime, 25% claimed that they would continue using it, while the rest conditioned their continued use on the exact amount to be charged and premium features. When asked about reasons for using Sawaal, 51% pointed at information access and knowledge about everyday matters including religion, health, news, education, and sports. Several users (11%) reported that they use it to connect to a virtual community of friends. Some users (1%) said that it is just a waste of time while 17% reported entertainment as the only benefit.

Knowledge Gaps and Retention

As discussed, we use Administrator Posted Questions to discover knowledge gaps, around specific topics, prevalent in the community and rephrased versions of these questions to measure knowledge retention.

Genre	#Qs	%Correct first attempts (N)	%Correct second attempts (N)	Δ	#Likes per Q	#Fwds per Q	#Cmts per Q
Smoking hazards	6	36.6 (1622)	70.3 (158)	33.7	3.8	24.8	4.2
Making a better society	4	69.1 (909)	93.5 (31)	24.4	4	29.3	2.5
Diabetes, heart, hypertension	24	43 (9166)	60.2 (762)	17.2	6.3	31.1	8.5
Mental health	12	44.4 (2869)	60.6 (251)	16.2	3.5	22	3.5
Food, water and hygiene	14	59.7 (4442)	73.6 (318)	13.9	4.6	44.2	5.1
Child safety and care	18	48.7 (5429)	62.4 (378)	13.7	3.9	27	4.1
Geography	2	74.1 (386)	87.5 (8)	13.4	3.5	6	2.5
Health, Hygiene, Nutrition and Medicine	8	62.9 (2141)	73.1 (160)	10.2	4.4	32.3	3.1
Dengue fever	12	50.8 (3054)	57.1 (168)	6.3	3.3	17.8	3.9
Special needs (physical and psychological)	8	46.3 (2464)	50 (174)	3.7	5.1	29.3	5.4
Dental health	2	69.7 (1073)	73 (141)	3.3	6.5	21	17.5
Breast cancer awareness	2	67.9 (863)	70.2 (57)	2.3	6	54	11.5
Maternal health	8	50.3 (2553)	51.7 (174)	1.4	4	14.8	6.3
General Knowledge	18	43.1 (5130)	42.7 (375)	-0.4	4.4	18.8	5.8
Population control and family planning	6	56.5 (2231)	53.6 (138)	-2.9	5.7	26.5	8
Environment and pollution	2	30.8 (663)	23.3 (43)	-7.5	6	26.5	5.5
First aid	20	47 (5761)	37.9 (425)	-9.1	3.9	27.3	4.5
Rules (traffic, civil) and regulations	8	50.1 (1650)	38.6 (88)	-11.5	2.9	14.6	2.4
Rights and constitution	10	63.9 (2660)	45.9 (157)	-18	3.4	36	4.5

Table 1: Administrator-posted Questions: By Genres

Administrator Posted Questions. 92 pairs of APQs were crafted to span various interesting and locally-relevant development-related topics. Both questions in each pair tested user's knowledge of the same concept and between them, we varied a subset of the question statement, answer choices or their ordering. To ensure reliability, all of the questions were taken from authoritative sources including *World Health Organization* [8], *Centers for Disease Control and Prevention* [2], *WebMD* [7], and *World Bank* [9]. Table 1 shows the topics and their distribution across the 184 questions (92 pairs). The questions are also marked with a perceived difficulty level by our content moderators (Table 2). This annotation is *not* based on the actual responses of the users. The questions are mostly easy or moderately difficult, with only four hard questions. Here are some example pairs of questions from each category. Correct answers are italicized.

Easy: *OQ*: How do you ensure the health of an expecting mom? (Answer choices: 1. Healthy, nutritious diet; 2. Timely medicines and vaccinations; 3. Reliable doctors; *4. All of these.*) *RQ*: How can you avoid complications during pregnancy? (Answer choices: 1. A good diet for the expecting mom; 2. Timely vaccinations and medicines; 3. Better medical facilities; *4. All of these.*)

Surprisingly, only 45% (out of 1,130) of users correctly answered the OQ, while 47% considered a good diet to be sufficient for the health of an expecting mother. The rephrased version was correctly answered by 63% of 89 users.

Medium: *OQ*: Where does a Dengue mosquito lay its eggs? (Answer choices: 1. Dirty stagnant water; 2. Clean stagnant water; *3. Both 1 and 2*; 4. In the hides of other animals.) *RQ*: Usually, which one of these is a common breeding place for Dengue mosquitoes? (Answer choices: 1. Unclean stagnant water; 2. Clean stagnant water; *3. Both 1 and 2*; 4. In the hides of other animals.)

This question was marked as moderately difficult due to a common misconception in the society (propagated via electronic media) that Dengue mosquitoes only lay their eggs in clean water. The OQ was attempted by 1,023 users but only 33% answered it correctly. The incorrect answers were split between 34% of users who chose dirty water and 32% who selected clean water (therefore the myth of clean water was not as uniformly prevalent as we had expected). The RQ was attempted by 41 users with 59% correct responses.

Question type	# Qs	%Correct first attempts (N)	%Correct second attempts (N)	Δ
Names	78	47.4 (21600)	59.8 (1417)	12.4
Opinions	48	51.6 (15875)	58.1 (1310)	6.5
Numeric	38	48.7 (9560)	52.3 (471)	3.6
Best Practices	20	55.8 (8031)	51.5 (718)	-4.3
Perceived Difficulty	**# Qs**	**%Correct first attempts (N)**	**%Correct second attempts (N)**	**Δ**
Hard	12	36.2 (3354)	57.6 (243)	21.4
Medium	88	45 (27358)	54.2 (2053)	9.2
Easy	84	57.7 (24354)	59.6 (1710)	1.9

Table 2: Administrator Questions: By Genre and Difficulty

Hard: *OQ*: What is "postpartum depression"? (Answer choices: 1. It's the same as other forms of depression; 2. It can follow a traumatic incident; *3. It can occur to a woman after she gives birth*; 4. It happens to a mother when she is separated from her child.) *RQ*: Which one of these psychological disorders could occur to a woman following childbirth? (Answer choices: *1. A specific kind of depression called postpartum depression*; 2. Women only experience physical changes to their bodies; 3. Insanity; 4. Brain tumors.)

Only 14% of 369 users attempted the OQ correctly. 35% thought that postpartum depression is the same as regular depression, while 28% thought that it occurs when a woman is separated from her child. The rephrased version was attempted by 21 users, 86% of whom answered it correctly.

Knowledge Gaps and Knowledge Retention. Table 1 shows the number of times the first and second questions in each pair were attempted (N) and the fraction of correct responses. As every question is presented to each user exactly once, the fraction of correct responses is equal to the fraction of users who correctly responded. The table also shows the number of likes, shares (forwards), and comments received by questions from each category. The first attempt of each question-pair reveals prevalent knowledge gaps. Table 1 shows that the most incorrectly attempted questions were regarding environment and pollution (30.8% out of 663 correct responses) and smoking hazards (36.6% correct responses out of 1,622). Less than half of all users exhibited correct knowledge regarding diabetes, heart, hypertension, healthy lifestyle, general knowledge, mental health, behavior towards people with special physical and psychological needs, first aid and child care. The most correctly answered questions belong to the categories of geography (74.1% correct responses), dental health (69.7% correct responses) and the making of a better society (69.1% correct responses). This last category comprised of questions regarding education being the right of every child and awareness regarding harms of littering. To

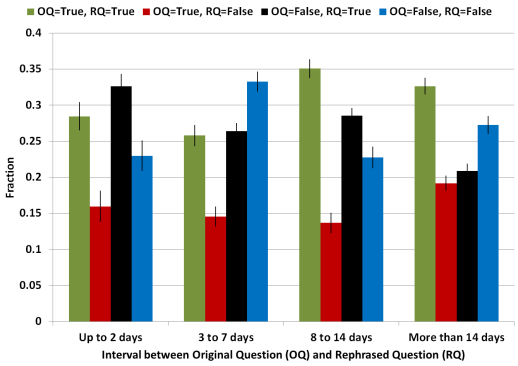

Figure 7: Fraction of correct first and second attempts

our surprise, users seem well aware regarding the perils of breast cancer.

As the correct answer is played to the users after their first attempt, the fraction of correct responses in the second attempt allows us to test not only knowledge retention but also possible misconceptions prevalent in the society. Such misconceptions could prevent knowledge absorption and participants may cling to their beliefs even after being informed of the correct reasons. For the set of genres we have in the system, we find that the highest improvement of correct responses occurs for smoking hazards (33.7% improvement), making of a better society (24.4%), diabetes, heart, hypertension, healthy lifestyle (17.2%) and mental health (16.2%). However, there are genres for which the fraction of correct responses actually falls for the second attempts. Rights and constitution, traffic laws, first aid, environment, and pollution are examples of such categories.

We find dental health, diabetes, heart, hypertension, and healthy lifestyle, breast cancer awareness, environment and pollution, population control and family planning to be the categories that were highly voted up and commented on. With 54 forwards per question, Breast cancer awareness seems to be the most shared category of questions. This is followed by food, water, and hygiene (44.2 forwards per question), and rights and constitution (36 forwards per question).

Table 2 shows the distribution of APQs by types and perceived difficulty. Questions involving names and opinions see more improvement between first and second attempt as compared to other types. The highest improvement (21%) was seen for hard questions, followed by moderately difficult questions (9%) and very little for easy questions.

In order to capture learning and *temporal knowledge retention* from the user's perspective, we compared the fraction of correct responses by users on original questions against corresponding rephrased questions using a paired t-test. The improvements are statistically significant for intervals of 2 days ($p<0.001$, $N=75$), 3-7 days ($p<0.001$, $N=70$), and 8-14 days ($p<0.001$, $N=72$). Although significant, the retention gradually decreases over the 14 days as shown in Figure 7 that compares fraction of correct responses to original and repeat questions. We do not find significant knowledge-retention ($p=0.15$, $N=81$) for intervals greater than 2 weeks (a catch-all bucket). This comparison also balances the effects of chance correct responses (probability = 0.25) between OQs and RQs.

Common Misconceptions and Interesting Observations: For this analysis, we only look at 163 questions that have been attempted 20 times or more. Interestingly, the most incorrectly answered question (0/21 correct responses) is: "Which one of these rights is every Pakistani entitled to?"(Answer choices: "1. Right to basic education, 2. Right to housing, 3. Right to higher education, 4. None of the above."). 95% of users thought that "Right to basic education" is one of the constitutional rights of every Pakistani. The correct response is "Right to housing", which is also a part of the slogan of a popular political party in Pakistan. Another interesting question was: "Which one of the following do not have a constitutional right to vote?" (Answer choices: "1. Men, 2. Women, 3. Children, 4. Old people"). This was answered correctly only twice out of 42 attempts. 64.2% users thought that old people, while 30.9% thought that men do not have the right to vote. No user thought that women do not have the right to vote. Out of 1,296, only 6.71% users knew that children need 12 vaccines up to the age of 5, while 63.27% think that they need 3 vaccines. 45.7% of 614 users knew that the best way to deal with a fussy baby is to leave the baby in a safe room for 10 minutes, 35% thought that they are better off whispering into her ear. Correct answers increased to 86% for the rephrased version of this question. 75% of 365 users seem to be well aware of the importance of mother's milk for healthy brain development of the baby. 74% of 354 users know that dengue is a mosquito-borne disease, 16.67% think that you can also get it by socializing with a dengue fever patient. 47.7% of 664 users think that use of air-conditioning can reduce global warming, 30.72% chose the correct response that use of such equipment should be reduced. 33% of 436 users think that babies born out of inter-caste marriages have a higher likelihood of having hereditary diseases. 91.89% of 555 users knew that children under 14 years of age should never be involved in labor and must concentrate on studies.

Information Preferences of the Community

We analyzed 1,000 randomly chosen CGQs, where each question had been attempted 50 times or more, to get an idea of what gets liked, disliked, commented on and shared based on a large number of attempts. Most of the questions (91.3%) contained actual content while 8.7% were empty (noise, silence). Of the 913 questions that had actual content, 98% were male-recorded while only 2% were recorded by females. Only 1% of

Genres	#Qs	%	#Attempts	%Correct Answers	#Likes per Q	#Reports per Q	#Dislikes per Q	#Comments per Q	#Fwds per Q
History	190	19	11740	54.2	0.9	1.1	0.2	0.6	6.6
Literature/Language	130	13	8165	50.4	0.9	0.8	0.3	0.8	5.7
Religion	108	10.8	6879	60.5	1.2	0.9	0.3	0.6	8.3
Sports	104	10.4	6415	44.8	0.9	1.5	0.4	0.7	5
Empty	74	7.4	4912	65	0.4	6.3	0.5	0.9	11.2
General Knowledge	70	7	4767	63.7	0.8	1.5	0.3	0.7	13.4
Geography	65	6.5	3962	49.8	0.9	1.8	0.3	0.8	4
Laws and Current Affairs	33	3.3	1990	55	0.5	1.8	0.4	1.1	4.3
Unclear	27	2.7	1747	47.5	1	3.6	0.3	1.1	9.1
Comedic	26	2.6	1571	54.9	0.6	4.7	0.5	1.9	1.9
Animals	23	2.3	1441	48.9	0.8	3	0.7	1.5	5.2
Community-related	21	2.1	1281	39.1	1.4	4.4	0.2	1.9	3.4
Personal/Self-related	21	2.1	1324	49.6	0.8	6.6	0.7	2.8	6.5
Everyday Science	19	1.9	1377	59.7	0.6	1	0.3	0.6	17.4
Noise	19	1.9	1134	60.3	0.2	6.6	0.3	0.9	3.4
Songs and music	17	1.7	1024	35.7	0.5	2.3	0.8	0.6	3.8
Quora like question	15	1.5	1026	65.8	1	2.8	0.4	1.3	12.7
Profane	11	1.1	707	45.4	0.5	8	1.1	2.6	3.3
Health, Nutrition and Medicine	8	0.8	502	49.2	0.9	3	0.3	1.5	4.3
Culture	6	0.6	354	50.6	1.2	0.8	0.7	1.3	1.3
Educational	6	0.6	345	72.2	1.3	2.3	0.3	0.3	2.3
Brand Awareness	3	0.3	202	49.5	0.3	3	0.3	1.3	10.3
Movies	3	0.3	170	32.9	0	8.3	0.7	1.3	1.3
Tech and IT	1	0.1	112	89.3	1	0	0	0	68
Grand Total	**1000**	**100**	**63147**	**54.1**	**0.9**	**2.2**	**0.4**	**0.9**	**7**

Table 3: Community-generated Questions: By Genre

recordings contained foul language or hate speech. 83.9% of the 1,000 community-generated questions were well-formed (in terms of the statement, answer choices and an option marked as correct), 1.1% were ambiguous, 2.9% were incomplete, 3.4% were not questions but announcements, poetry, songs or comments. The answer choices marked correct were well distributed over the four options with some bias towards option 1 (35.10%, 18.3%, 24.4%, 22.2% for options 1 – 4 respectively). Users asked about names in 51.5% of all questions (e.g. "Which one of these cities is the capital of USA?"), quantities (speed, distance, volume, amount, height etc.) in 22.5%, opinions in 8.4%, and dates in 5.1% of all questions.

Our moderators marked the genre for the CGQs (Table 3). Sports-related questions were mostly about cricket (players, teams, history). 67% of all questions contained religious content. Geographical questions were about atmosphere, weather, seasons, and landmarks. Categorized as "songs and music", people recorded songs and tunes from other devices and asked questions about them. Literature related questions were about poetry, books, authors, words, phrases. Open-ended questions (similar to questions posted on Quora) asked for opinions of others, e.g: "how does one get good marks in exams?". "Community-related" questions were messages for other users of Sawaal e.g. "What is the name of the user who cheats to get high scores?". "Comedic questions" were jokes and puzzles posted for entertainment e.g. "Why does a donkey have two ears?". "Educational" questions were related to learning e.g "How many seconds are there in a day?".

The questions that were answered correctly 50% of times or more were about technology (89%), education (72%), general knowledge (63.7%), religion (60.5%), and open-ended (65.8%). Community-related, educational, religious and cultural questions are among the most liked ones. Questions mostly reported for abuse include empty and noisy ones and questions about movies, profanity and personality of the users who posted them. Profane questions are also the ones with highest number of dislikes and comments per question. Several users in the community got into heated arguments (via comments) with people who post profane questions.

Some Interesting Uses of Sawaal: Users compete hard and take a lot of pride in their scores. Some also found ways to game the system to achieve high scores. Initial methods involved forwarding a question to one's own number. When we banned self-forwarding, users started forwarding questions to their friends who forward these back to them. Another trick involves sending an empty question to friends who apparently already know which option is marked correct. As a result, we see questions marked empty or noise being forwarded a lot. Frustrated by sudden high scores of their opponents, some people started complaining about this in audio comments, user feedback and posted questions.

A radio host posted informative questions from radio shows. Some started "themed, hosted quiz shows" that ask a series of questions about various topics (current affairs, sports, science, religion, language, geography). A user taught Arabic by recording explanations of terms and concepts followed by questions. A musician played music and asked others to identify the tune and instrument played. Incorrect/incomplete questions were pointed out and often corrected in response by the people who posted them. In comments, users debated posting ethics. Several users improved their posting habits in response to such feedback.

We performed an analysis of comments and found 34.6% empty files in a random sample of 500. 95% of the remaining 327 files were male-recorded. As compared to questions, a larger fraction of comments (30%) contained foul language and heated arguments around posted questions, probably because comments are not moderated. Most of the comments were regarding posted questions and contained criticism (13.4%), praise (14.5%), corrections (8.7%) and suggestions for improvement of the questions (13.4%). Clearly, comments were used as a discussion forum around posted questions.

Analysis of User Feedback

Sawaal allows users to provide feedback and suggestions in the form of unconstrained speech. Based on a sample of 500 feedback recordings, there were 42.4% empty files (comprising noise and silence), while the remaining 57.6% (N=288) had actual content. Of these, 88% of the recordings were by male users. A large fraction of users praised the service profusely (36% of all recordings) and said that it helps them improve their general knowledge and provides them with information about everyday matters and health. In 75% of recordings, users suggested improvements to Sawaal like the removal of empty, incoherent and inappropriate

recordings. They also suggested that there should be a limit on the number of questions recorded by any one user.

5 DISCUSSION AND FUTURE WORK

We have suggested a new method of designing information campaigns in this paper. We show that the use of voice-based quizzes comprising multiple choice questions can lead to effective information dissemination. Gamification, competition and active-collaborative learning are fostered through quizzes, with an additional benefit of being able to measure retention of the disseminated information among the target audience. Our work shows that social connectivity, the spirit of competition and the agency to enrich the knowledge of the community are effective motivators for making an information service engaging and attractive. Users not only contribute content enthusiastically but also thoroughly engage with all aspects of the interface and even take an active part in spreading the service to others. This leads to an organic spread of the information service in the society hence making explicit advertisement of the service, redundant. In our deployment, the service reached mostly young and low-literate men including a lot of school students.

The goals of an information campaign can dictate the administrator-posted questions. We have shown that user responses could reveal knowledge gaps. Once they find out the answers to the questions, the rephrased version presented to them later allows measuring temporal, subject-wise and difficulty-wise knowledge retention. For Sawaal, the retention remained significant for up to two weeks. As a significant fraction of users remains engaged with Sawaal for several weeks, they can be offered the rephrased questions repeatedly to further improve retention. In our example deployment, we only created one rephrased version of each question and repeated each content type no more than twice (original question and rephrased version). For campaigns aiming for greater assimilation of the delivered knowledge, the repetitions and rephrased versions could be increased.

Our work also shows that community-generated questions not only help engage users, promote collaboration, and enhance ownership of the contributed information but also reveal the information needs of the society. Users' ability to contribute questions and challenge others to play for high scores, allows them to come up with creative uses of the service like teaching courses and designing radio-style quiz "shows". Our initial intervention had questions on diverse genres to allow us to draw generalizable conclusions. Our current focus is on using the same technique for more targeted interventions.

Finally, we believe that our work can be adopted and extended in a variety of ways to fit the needs of specific campaigns. In the future we plan to test the impact of variables like playing more detailed answers to administrator posted questions, may be endorsed by people of social influence (like celebrities), more repetitions of questions, and more flavors of rephrasing of questions. A skip option would be added to answer choices so that users are not forced to choose a random answer when they do not know the correct response. We also plan to incorporate auto-detection of foul language and empty recordings.

6 CONCLUSION

Our work shows that voice-based quizzes provide a viable way to measure knowledge-impact of information dissemination campaigns targeted towards under-connected populations, at a large scale and in real-time. Our service, Sawaal, provides a way to measure topic-specific existing knowledge gaps and misconceptions in the society, to disseminate correct information, and finally to measure the extent and duration of knowledge retained. Community-contributed quiz content and an ability to play against your friends for high scores, leads to more inclusion and ownership, active-collaborative learning and a spirit of competition among the users. The service spreads organically among the target audience, receives enthusiastic user response, and successfully retains a significant fraction of the users for several weeks. It reaches otherwise under-connected people including low-literates, visually impaired, tech naïve and school students in rural and remote locations.

7 LIMITATIONS

Instruments like Sawaal can only measure knowledge gaps and misconceptions prevalent in local communities pertaining to specific topics and subjects that the researchers already expect or can hypothesize about. It is often challenging for 'outsiders' to interpret local knowledge as there might be differences in interpretation of the questions and responses between the target community and intervention designers [17, 23, 28, 54, 55, 58]. In terms of knowledge retention, our results only prove significant effects for up to two weeks. We did not have enough data to extend our claims beyond this duration. Another limitation is the simplifying assumption based on literature that there is a one-on-one mapping between users and phone numbers, which is only true for a subset of them. Finally, as this was the first exploratory project of its nature, we did not have a baseline to compare against or a control to isolate confounding factors.

ACKNOWLEDGMENTS

We thank ITU for supporting this work, our reviewers for their insightful feedback, our RAs Hira Ejaz, Haris Bin Zia, Abdullah Sabri, and Mukhtar Ahmad for spending hours annotating data, moderating posts and performing surveys.

REFERENCES

[1] 2017. Capacity Plus. (16 Sept. 2017). https://www.k4health.org/sites/default/files/ivr-report-final.pdf

[2] 2017. Centers for Disease Control and Prevention. https://www.cdc.gov. (2017).

[3] 2017. Duolingo. (14 Sept. 2017). https://itunes.apple.com/us/app/duolingo-learn-spanish-french-and-more/id570060128?mt=8

[4] 2017. Gram Vaani. http://www.gramvaani.org/. (2017).

[5] 2017. Magoosh. (14 Sept. 2017). https://itunes.apple.com/us/developer/magoosh/id522118006

[6] 2017. Sporcle. (14 Sept. 2017). https://www.sporcle.com/

[7] 2017. WebMD - Better information, Better health. http://www.webmd.com. (2017).

[8] 2017. WHO | World Health Organization. http://www.who.int. (2017).

[9] 2017. World Bank Group - International Development, Poverty, & Sustainability. http://www.worldbank.org/. (2017).

[10] Mohammed Arif, Charles Egbu, Ola Alom, and Malik MA Khalfan. 2009. Measuring knowledge retention: a case study of a construction consultancy in the UAE. *Engineering, Construction and Architectural Management* 16, 1 (2009), 92–108.

[11] Douglas S Bell, Charles E Harless, Jerilyn K Higa, Elizabeth L Bjork, Robert A Bjork, Mohsen Bazargan, and Carol M Mangione. 2008. Knowledge retention after an online tutorial: a randomized educational experiment among resident physicians. *Journal of general internal medicine* 23, 8 (2008), 1164–1171.

[12] Carl A Benware and Edward L Deci. 1984. Quality of Learning With an Active Versus Passive Motivational Set. *American Educational Research Journal* 21, 4 (1984), 755–765. DOI: http://dx.doi.org/10.3102/00028312021004755

[13] Anke Berns, Antonio Gonzalez-Pardo, and David Camacho. 2013. Game-like language learning in 3-D virtual environments. *Computers & Education* 60, 1 (2013), 210–220.

[14] Nicola J Bidwell and Masbulele Jay Siya. 2013. Situating asynchronous voice in rural Africa. In *IFIP Conference on Human-Computer Interaction*. Springer, 36–53.

[15] Lorena Blasco-Arcas, Isabel Buil, Blanca Hernández-Ortega, and F Javier Sese. 2013. Using clickers in class. The role of interactivity, active collaborative learning and engagement in learning performance. *Computers & Education* 62 (2013), 102–110.

[16] Deana Brown, Gary Marsden, and Ulrike Rivett. 2012. WATER alert!: using mobile phones to improve community perspective on drinking water quality in South Africa. In *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development*. ACM, 230–240.

[17] John Seely Brown and Paul Duguid. 2001. Knowledge and organization: A social-practice perspective. *Organization science* 12, 2 (2001), 198–213.

[18] Sebastien Cuendet, Indrani Medhi, Kalika Bali, and Edward Cutrell. 2013. VideoKheti: Making video content accessible to low-literate and novice users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2833–2842.

[19] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*. ACM, 9–15.

[20] Robert Godwin-Jones. 2014. Games in language learning: Opportunities and challenges. (2014).

[21] Mohamed Gulaid and Aditya Vashistha. 2013. Ila Dhageyso: an interactive voice forum to foster transparent governance in Somaliland. In *Proceedings of the Sixth International Conference on Information and Communications Technologies and Development: Notes-Volume 2*. ACM, 41–44.

[22] Paul Haidet, Robert O Morgan, Kimberly O'malley, Betty Jeanne Moran, and Boyd F Richards. 2004. A controlled trial of active versus passive learning strategies in a large group setting. *Advances in Health Sciences Education* 9, 1 (2004), 15–27.

[23] Tim Ingold. 2011. *Being alive: Essays on movement, knowledge and description*. Routledge.

[24] Matthew Kam, Aishvarya Agarwal, Anuj Kumar, Siddhartha Lal, Akhil Mathur, Anuj Tewari, and John Canny. 2008. Designing e-learning games for rural children in India: a format for balancing learning with fun. In *Proceedings of the 7th ACM conference on Designing interactive systems*. ACM, 58–67.

[25] Matthew Kam, Divya Ramachandran, Varun Devanathan, Anuj Tewari, and John Canny. 2007. Localized iterative design for language learning in underdeveloped regions: the PACE framework. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1097–1106.

[26] Zahir Koradia, Piyush Aggarwal, Aaditeshwar Seth, and Gaurav Luthra. 2013. Gurgaon idol: a singing competition over community radio and IVRS. In *Proceedings of the 3rd ACM Symposium on Computing for Development*. ACM, 6.

[27] Anuj Kumar, Pooja Reddy, Anuj Tewari, Rajat Agrawal, and Matthew Kam. 2012. Improving literacy in developing countries using speech recognition-supported games on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1149–1158.

[28] Jean Lave. 2009. The practice of learning. *Contemporary theories of learning: Learning theoristsâĂę in their own words* (2009), 200–208.

[29] Adam Lerer, Molly Ward, and Saman Amarasinghe. 2010. Evaluation of IVR data collection UIs for untrained rural users. In *Proceedings of the First ACM Symposium on Computing for Development*. ACM, 2.

[30] Alexander Lex, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister. 2014. UpSet: Visualization of Intersecting Sets. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '14)* 20, 12 (2014), 1983–1992.

[31] Klaus Libertus and Amy Needham. 2010. Teach to reach: The effects of active vs. passive reaching experiences on action and perception. *Vision research* 50, 24 (2010), 2750–2757.

[32] Mywish K Maredia, Byron Reyes, Malick N Ba, Clementine L Dabire, Barry Pittendrigh, and Julia Bello-Bravo. 2017. Can mobile phone-based animated videos induce learning and technology adoption among low-literate farmers? A field experiment in Burkina Faso. *Information Technology for Development* (2017), 1–32.

[33] J Patrick McCarthy and Liam Anderson. 2000. Active learning techniques versus traditional teaching styles: Two experiments from history and political science. *Innovative higher education* 24, 4 (2000), 279–294.

[34] Indrani Medhi, Raghu S Menon, Edward Cutrell, and Kentaro Toyama. 2012. Correlation between limited education and transfer of learning. *Information Technologies & International Development* 8, 2 (2012), pp–51.

[35] Indrani Medhi, S Raghu Menon, Edward Cutrell, and Kentaro Toyama. 2010. Beyond strict illiteracy: abstracted learning among low-literate users. In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*. ACM, 23.

[36] Indrani Medhi, Somani Patnaik, Emma Brunskill, SN Gautama, William Thies, and Kentaro Toyama. 2011. Designing mobile interfaces for novice and low-literacy users. *ACM Transactions on Computer-Human Interaction (TOCHI)* 18, 1 (2011), 2.

[37] Indrani Medhi-Thies, Pedro Ferreira, Nakull Gupta, Jacki O'Neill, and Edward Cutrell. 2015. KrishiPustak: a social networking system for low-literate farmers. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1670–1681.

[38] Maletsabisa Molapo, Melissa Densmore, and Limpho Morie. 2016. Apps and Skits: Enabling New Forms of Village-To-Clinic Feedback for Rural

Health Education. In *Proceedings of the 7th Annual Symposium on Computing for Development*. ACM, 10.

[39] Preeti Mudliar, Jonathan Donner, and William Thies. 2012. Emergent practices around CGNet Swara, voice forum for citizen journalism in rural India. In *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development*. ACM, 159–168.

[40] Michelle L Munro, Jody R Lori, Carol J Boyd, and Pamela Andreatta. 2014. Knowledge and skill retention of a mobile phone data collection protocol in rural Liberia. *Journal of midwifery & women's health* 59, 2 (2014), 176–183.

[41] Manuel Palomo-Duarte, Anke Berns, Alberto Cejas, Juan Manuel Dodero, Juan Antonio Caballero, and Iván Ruiz-Rube. 2016. Assessing foreign language learning through mobile game-based learning environments. *International Journal of Human Capital and Information Technology Professionals (IJHCITP)* 7, 2 (2016), 53–67.

[42] Neil Patel, Sheetal Agarwal, Nitendra Rajput, Amit Nanavati, Paresh Dave, and Tapan S Parikh. 2009. A comparative study of speech and dialed input voice interfaces in rural India. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 51–54.

[43] Neil Patel, Deepti Chittamuru, Anupam Jain, Paresh Dave, and Tapan S Parikh. 2010. Avaaj otalo: a field study of an interactive voice forum for small farmers in rural india. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 733–742.

[44] Udai Singh Pawar, Joyojeet Pal, Rahul Gupta, and Kentaro Toyama. 2007. Multiple mice for retention tasks in disadvantaged schools. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1581–1590.

[45] Trevor Perrier, Abhigyan Kaustubh, Abhishek Gupta, and Richard Anderson. 2013. Questioning feedback: improving public health messaging. In *Proceedings of the Sixth International Conference on Information and Communications Technologies and Development: Notes-Volume 2*. ACM, 112–115.

[46] Agha Ali Raza, Awais Athar, Shan Randhawa, Zain Tariq, Muhammad Bilal Saleem, Haris-Bin Zia, Umar Saif, and Roni Rosenfeld. 2018. Rapid Collection of Spontaneous Speech Corpora Using Telephonic Community Forums. *Proc. Interspeech 2018* (2018), 1021–1025.

[47] Agha Ali Raza, Mansoor Pervaiz, Christina Milo, Samia Razaq, Guy Alster, Jahanzeb Sherwani, Umar Saif, and Roni Rosenfeld. 2012. Viral entertainment as a vehicle for disseminating speech-based services to low-literate users. In *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development*. ACM, 350–359.

[48] Agha Ali Raza, Samia Razaq, Amna Raja, Rizwan Naru, Ali Gibran, Abdullah Sabri, Haroon Niaz, Muhammad Bilal Saleem, and Umar Saif. 2016. Real-time Automated Surveys among Low-literate Masses using Voice-based Telephone Services. In *Proceedings of the 7th Annual Symposium on Computing for Development*. ACM, 17.

[49] Agha Ali Raza, Bilal Saleem, Shan Randhawa, Zain Tariq, Awais Athar, Umar Saif, and Roni Rosenfeld. 2018. Baang: A Viral Speech-based Social Platform for Under-Connected Populations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 643.

[50] Agha Ali Raza, Farhan Ul Haq, Zain Tariq, Mansoor Pervaiz, Samia Razaq, Umar Saif, and Roni Rosenfeld. 2013. Job opportunities through entertainment: Virally spread speech-based services for low-literate users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2803–2812.

[51] Agha Ali Raza, Farhan Ul Haq, Zain Tariq, Mansoor Pervaiz, Samia Razaq, Umar Saif, and Roni Rosenfeld. 2013. Job opportunities through entertainment: Virally spread speech-based services for low-literate users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2803–2812.

[52] Robert Roche, Eric Hladilek, and Scott Reid. 2006. Disaster recovery virtual roll call and recovery management system. (April 11 2006). US Patent 7,026,925.

[53] Bruce Rocheleau and Liangfu Wu. 2005. E-Government and financial transactions: Potential versus reality. *The Electronic Journal of e-Government* 3, 4 (2005), 219–230.

[54] Edward Said. 1994. Orientalism. 1979. *New York: Vintage* (1994).

[55] Edward W Said. 1985. Orientalism reconsidered. *Race & class* 27, 2 (1985), 1–15.

[56] Jahanzeb Sherwani, Nosheen Ali, Sarwat Mirza, Anjum Fatma, Yousuf Memon, Mehtab Karim, Rahul Tongia, and Roni Rosenfeld. 2007. Healthline: Speech-based access to health information by low-literate users. In *Information and Communication Technologies and Development, 2007. ICTD 2007. International Conference on*. IEEE, 1–9.

[57] Mel Silberman. 1996. *Active Learning: 101 Strategies To Teach Any Subject*. ERIC.

[58] Gayatri Chakravorty Spivak. 2012. *In other worlds: Essays in cultural politics*. Routledge.

[59] C-H Su and C-H Cheng. 2015. A mobile gamification learning system for improving the learning motivation and achievements. *Journal of Computer Assisted Learning* 31, 3 (2015), 268–286.

[60] Abhishek Behl23 Maneesha Vashishtha24. 2014. Innovative IVR system for farmers: enhancing ICT adoption. (2014).

[61] Aditya Vashistha, Edward Cutrell, Gaetano Borriello, and William Thies. 2015. Sangeet swara: A community-moderated voice forum in rural india. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 417–426.

[62] Aditya Vashistha, Edward Cutrell, Nicola Dell, and Richard Anderson. 2015. Social media platforms for low-income blind people in india. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. ACM, 259–272.

[63] Aditya Vashistha, Pooja Sethi, and Richard Anderson. 2017. Respeak: A Voice-based, Crowd-powered Speech Transcription System. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 1855–1866.

[64] Aditya Vashistha, Pooja Sethi, and Richard Anderson. 2018. BSpeak: An Accessible Voice-based Crowdsourcing Marketplace for Low-Income Blind People. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 57.

[65] Jerome White, Mayuri Duggirala, Krishna Kummamuru, and Saurabh Srivastava. 2012. Designing a voice-based employment exchange for rural India. In *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development*. ACM, 367–373.

[66] Robert A Wisher, Christina K Curnow, and Robert J Seidel. 2001. Knowledge retention as a latent outcome measure in distance learning. *American Journal of Distance Education* 15, 3 (2001), 20–35.

[67] Nikolas Wolfe, Juneki Hong, Agha Ali Raza, Bhiksha Raj, and Roni Rosenfeld. 2015. Rapid development of public health education systems in low-literacy multilingual environments: Combating ebola through voice messaging. In *ISCA Special Interest Group on Speech and Language Technology in Education (SLaTE)*. INTERSPEECH.

[68] Lincoln Wood and Torsten Reiners. 2012. Gamification in logistics and supply chain education: Extending active learning. *internat technologies & society 2012* (2012), 101–108.