Karamad: A Voice-based Crowdsourcing Platform for Underserved Populations

Shan M Randhawa
smr693@nyu.edu
New York University, Abu Dhabi
United Arab Emirates

Tallal Ahmad
tallal.ahmad@lums.edu.pk
Lahore University of Management Sciences (LUMS)
Pakistan

Jay Chen
jchen@icsi.berkeley.edu
International Computer Science Institute (ICSI)
Berkeley, California, USA

Agha Ali Raza
agha.ali.raza@lums.edu.pk
Lahore University of Management Sciences (LUMS)
Lahore, Pakistan

ABSTRACT

Crowdsourcing enables the completion of large-scale and hard-to-automate tasks, while allowing people to earn money. However, 3.6 billion people – a workforce comprising 46.4% of the world population – who could benefit most from this source of income lack the access and literacy to use computers, smartphones, and the internet. In this paper we present *Karamad*, a voice-based crowdsourcing platform that allows workers in low-resource regions to complete crowd work using low-end phones and receive payments as mobile airtime balance. We explore the usefulness, scalability, and sustainability of Karamad in Pakistan through a 6-month deployment. Without any advertising, training, or airtime subsidies, Karamad organically engaged 725 workers who completed 3,939 tasks (involving 43,006 components) including translations, dataset generation, and surveys on demographics, accessibility, disability, health, employment, and literacy. Collectively, the workers produced a valuable service market for potential customers and included female, unemployed, non-literate, and blind users.

CCS CONCEPTS

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

KEYWORDS

HCI4D; ICT4D; crowdsourcing; telephonic surveys; underserved; Interactive Voice Response; IVR; Pakistan; low-literate; mobile phone; income; crowd work; voice; speech; telephone; cellphone.

ACM Reference Format:
Shan M Randhawa, Tallal Ahmad, Jay Chen, and Agha Ali Raza. 2021. Karamad: A Voice-based Crowdsourcing Platform for Underserved Populations. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3411764.3445417

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '21, May 8–13, 2021, Yokohama, Japan

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445417

1 INTRODUCTION

Crowdsourcing has been a highly productive area for HCI research for over a decade. Mainstream platforms such as Amazon Mechanical Turk [1] provide a marketplace where requesters issue tasks to workers that, upon completion, produce value in the form of tagging, translation, surveys etc. For the workers, completing such tasks can be fun and also serve as a source of supplementary income yielding roughly $2 per hour [24]. Unfortunately, despite over a decade of HCI research in crowdsourcing, crowd work is still not generally accessible to many underserved communities.

Most publicly available crowdsourcing platforms are designed with the assumption that workers have access to, and are able to use, a computer or smartphone and the internet. According to the GSMA, 1.8 billion people worldwide have access to feature phones, but not smartphones [22]. Around 1.4 billion people have a mobile service subscription but not internet access [22]. Another 750 million youth and adults in the world are illiterate [33], and 2.2 billion people have some type of visual impairment that makes the use of visual interfaces difficult or impractical for them. These communities also overlap with the 736 million people who live below the poverty line, earn $2 per day [5], and could benefit most from the income generated by crowd work.

Past research on crowdsourcing considered ways to address some of the salient accessibility, usability, and equity challenges that affect crowdsourcing platforms [21]. Recently, researchers focusing on inclusion have explored voice-based platforms as a promising way forward due to the accessibility, usability, and ubiquity of speech interfaces [49, 50, 61, 63]. For example, BSpeak [66] focuses on assessing the ability of blind people in India to perform transcription tasks and its financial feasibility and ReCall [63] evaluates the accuracy of audio transcription tasks performed by low-income users on feature phones.

Inspired by this literature we developed, Karamad[1], a voice-based crowdsourcing platform to explore the potential of crowd work in underserved communities.

To assess the types of possible voice-based tasks, we brainstormed the potential use cases by surveying existing data collection efforts for different purposes. To demonstrate that the use cases can be implemented in Karamad, we replicated existing studies and

[1]Karamad means "useful" in Urdu. It is a play on words as "kar" also means "work" and "amad" means "arrival", so it can translate to "Work has arrived!".

created our own tasks. To evaluate whether tasks can be realistically completed by workers in a scalable and sustained manner, we deployed Karamad publicly in Pakistan for 6 months.

This paper makes two main contributions. First, we explore the design of Karamad, a general purpose voice-based crowdsourcing platform for low-resource workers that does not require users to have access to a smartphone or the internet. Second, this paper presents lessons learned from a 6 month long in-situ deployment in Pakistan without airtime subsidies, user-training, or advertisement. In this paper, we do not focus on the results of the tasks themselves other than as a means to assess Karamad in terms of the workers who used our system, the ability of the workers to complete tasks, and Karamad's sustainability. Instead, we focus on a quantitative analysis of worker behavior on our platform and present qualitative findings from 35 telephone interviews.

Overall, we find that Karamad can be used for surveys, translation, corpus collection, and other tasks that were previously conducted via in-person visits. Karamad's protective measures and financial incentives were sufficient to attract workers to use the platform and to perform tasks over a sustained period of 6 months. During our study, Karamad spread organically from 0 to 725 workers and was limited in scale only by the limitations of the existing payment gateways and our research budget. Karamad's workers completed 3,939 tasks involving 43,006 components during the study and included 11% female, 61% unemployed, 40% non-literate, and 85% blind users based on 313 self-selected demographics survey responses. Of our 35 interview respondents, 57% reported not having any income or opportunities for income, and relying instead on relatives for assistance.

2 RELATED WORK

2.1 Crowd Work

Crowdsourcing is a technology that enables the supply of human labor to be leveraged to meet the demand for tasks that require human intelligence. Initially, crowdsourcing focused on cognitive tasks, but mainstream crowdsourcing platforms like Amazon's Mechanical Turk (MTurk) [1] have rapidly expanded to more complex and even creative work [43]. The growth of crowdsourcing has led to significant shifts in the landscape of employment and has contributed to the new "gig economy" where independent labor is hired for short-term jobs.

This major change to the nature and future of work often comes at an invisible cost to workers [21]. Hard-won progress in labor regulation including the establishment of a minimum wage, enforcement of humane working hours, and prevention of child labor are effectively non-existent in this new and under-regulated market. Irani and Silberman recognized the lack of transparency and recourse afforded to the crowd workers and developed a tool, Turkopticon [26], to help protect MTurk workers from being exploited by task requesters and the lack of transparency of the platform itself. Savage et al. [55] found that various strategies and tools including Turkopticon are used by online communities in attempts to reduce the amount of unpaid work and to earn more money. Silberman et al. [59] proposed guidelines for the treatment of crowd workers that included the recommendation of paying at least minimum wage. However, more effort is needed to improve fairness for workers and reduce their exploitation by crowdsourcing platforms [24].

Despite the shortcomings of crowd work, it is still a potentially important source of income for workers, particularly for people who cannot find formal employment. Studies on the demographics of MTurk are varied depending on the methodology. Difallah et al. [10] found that 80% of MTurk workers are from the United States, and the majority of the remaining workers are from India, whereas Pavlik et al. [46] found that more workers were from India based on the geo-location of their IP addresses. However, while these statistics show the ongoing usage of MTurk, they are not representative of the non-users who may be interested in crowd work. For people living in poor countries or in rural regions, MTurk and most other platforms are not accessible because it requires people to have English language skills, access to computers, and the internet. Attempts to make MTurk more accessible to low-income workers in India by Khanna et al. [29] found that simplified user interfaces, simplified task instructions, and language localization are absolutely necessary.

2.2 Inclusion and Developing Contexts

Researchers are interested in making crowdsourcing more inclusive and accessible to developing areas due to its potential for increasing the income of the poor, but large platforms and organizations that need crowd work have largely overlooked rural workers due to their lack of technology and their uncertain capabilities. Early crowdsourcing platforms trying to bridge the gap for developing areas like mClerk [23] and MobileWorks [40] were therefore mobile-based and pioneered the inclusion of rural low-income workers by having them transcribe images sent to their phones via SMS and mobile web-based applications. txtEagle [12] is one of the very first mobile crowdsourcing platforms for developing regions that gathered SMS-based survey responses. These systems demonstrated the feasibility of crowd work by rural and low-income populations, but were limited in that they expected crowd workers to have reading and typing skills or English proficiency.

Some modern crowdsourcing platforms such as Samasource [54] have extended crowd work to developing areas, but Samasource still requires workers to work from cybercafes and to undergo extensive training in English. Other than the access barriers preventing the inclusion of rural or low-income workers, Calvo et al. [7] also found that access to crowd work could be particularly useful for workers with disabilities. Hara et al. [25] explored how crowd work could be made accessible to workers with autism spectrum disorder. Vashistha et al. [66] have made further explorations into making crowd work accessible to visually impaired workers in India.

In parallel with developments in crowdsourcing, Interactive Voice Response (IVR) systems have increased in maturity and presence in developing contexts. Although IVRs are more constrained than an internet-enabled computer, IVR-based systems have far greater reach in rural developing contexts and their capabilities closely complement the competence of low-income, rural, and visually impaired users. Consequently, IVR-based platforms have generated increased interest over the past decade in developing contexts and have been aggressively deployed in developing regions to enable a wide range of services including social media [49, 62], education [50, 61], agriculture [45], healthcare [41], citizen journalism [35], job search [51], and data collection [32].

Our work is situated at the intersection of crowdsourcing and speech-based services for developing regions. The closest literature to Karamad include Respeak [65], BSpeak [66], Recall [63], and work by Chopra et al. [9]. Respeak is a voice-based system built as an Android application that employs low-income users to transcribe audio files in English and Hindi. BSpeak is a similar platform that allows blind users to transcript audio via speech and automatic-speech recognition. Recall is similar to Respeak, but uses an IVR rather than an Android application, and focuses on evaluating the error rate of transcription tasks over an 8kHz voice channel. Each of these systems leverages users' speaking and listening abilities, and Respeak and Bspeak also require users to have access to smartphones. Recently, Chopra et al. [9] explored an Android-based crowdsourcing platform for allowing low-income users to perform digitization tasks in India in a two-week study. Karamad differs from these systems in that it is a general crowdsourcing platform that enables data collection tasks beyond audio transcription or digitization. Furthermore, Karamad runs on a feature phone without the need of and internet access. Our work focuses on the in-situ scalability and long-term sustainability of crowd work rather than controlled short-term experiments.

3 KARAMAD

Karamad is a voice-based crowdsourcing platform complete with task creation, task issuance, payment, and analytics, that allows requesters to create tasks for workers. In this section we discuss the design of Karamad, its capabilities, and its interfaces for workers and requesters. As with existing crowdsourcing platforms, the overarching goal of Karamad is to allow requesters to hire crowd workers to complete tasks. Additionally, Karamad aims to include underserved populations (marginalized, under-connected, low-literate, poor) into crowd work as a means to supplement their income. In this paper we focus on presenting and evaluating the worker-facing aspects of Karamad.

3.1 Design

We use an IVR to interact with our workers who are typically low-literate, non-tech-savvy, poor, or visually impaired. The interaction modality of an IVR is more constrained compared to a computer or smartphone with internet and multimedia capabilities. Furthermore, the capacities of the target demographics are also expected to be different compared to educated workers in the US or India.

To map out the potential of the task types that are both realizable through an IVR and by our workers, we considered the types of tasks available on Mturk and we excluded tasks that involved images or videos due to the limitations of the voice-based modality and tasks that require high bandwidth or cognitive load that are not likely to be suitable for a phone call (e.g. sifting through large documents). Compared to the types of tasks on MTurk [1], image or video processing tasks are not feasible via IVR interfaces. Based on the literature [8, 20, 44, 49, 50, 60, 61, 63], information gathering and data processing tasks such as conducting surveys and gathering speech recordings are the most viable for our workers and in demand. We focused on supporting these two types of tasks.

3.2 Implementation

Karamad is composed of three main components: A web portal, an IVR platform, and a payment processor. The web portal is used by requesters to design, deploy, and monitor tasks. The IVR platform is used by workers to find and complete tasks to earn compensation in the form of mobile top-ups (i.e. mobile airtime payments). The payment processor keeps track of scheduled compensation for workers and performs airtime transfers. We implemented a survey-task web-portal using Laravel to allow requesters to define tasks. The IVR platform was implemented using FreeSWITCH [18], an open-source telephony platform, supported with a PRI (Primary Rate Interface) phone line.

Unlike most prior works in developing areas, we do not employ any airtime subsidy mechanisms such as toll-free numbers or missed-calls. Instead, the users bear the airtime cost themselves and we pay workers with mobile airtime after tasks are completed. As a result, we were careful throughout our implementation and deployment to avoid having users waste their airtime and to ensure that they are compensated fairly.

To prevent users from wasting airtime when they are not eligible for attempting any available tasks, we created a pre-screening mechanism using *early media* in our IVR. Early media (also called the ring-back-tone) is a feature of telephony systems that allows sending audio to a caller prior to answering their call. Usually this mechanism is used to play appropriate tones and messages to inform the caller that the recipient's phone is ringing or busy. Since this takes place before the call is answered, the caller is not charged for this "pre-answer" period.

To pay our workers, we considered 5 different airtime payment portals: Easypaisa [13], Jazzcash [28], Ding [11], Recharge [52], and Meezan [3]. Each payment portal had various restrictions that made them inconvenient for frequent small payments or had high overheads (up to 37%). We use Meezan Internet Banking as our payment gateway because it only charges 12.5% tax and has a web portal with arbitrary top-up amounts. We accumulate payments of each worker and pay them once every few days as Meezan only allows 5 transactions per network per day, and 20 transactions per day of amounts between PKR 40 and PKR 2,500 (PKR 167 = 1 USD). Most mobile payment gateways require the mobile carrier to be on the same network to deliver a mobile top-up. Our payment processor used an HLR (Home Location Registry) lookup service [34] to retrieve mobile carrier information and Selenium [57] to automate the payment process.

3.3 Capabilities

To map out the set of potentially useful task types, we examined the literature for works that have conducted IVR surveys or quizzes (e.g. [8, 20, 44, 50, 60, 61]). We then implemented sufficient capabilities in Karamad to accommodate all tasks that we intended to deploy. Karamad currently supports Yes-No (binary) questions, multiple choice questions (MCQ), short and long answer questions. Binary questions and multiple choice questions require a single key press as input. Short and long answer questions require free-form 30 and 60 second voice recordings. We did not implement longer speech recordings or multiple digit inputs in Karamad that have been used in the past by IVR-based services [8, 41, 49, 50]. We made

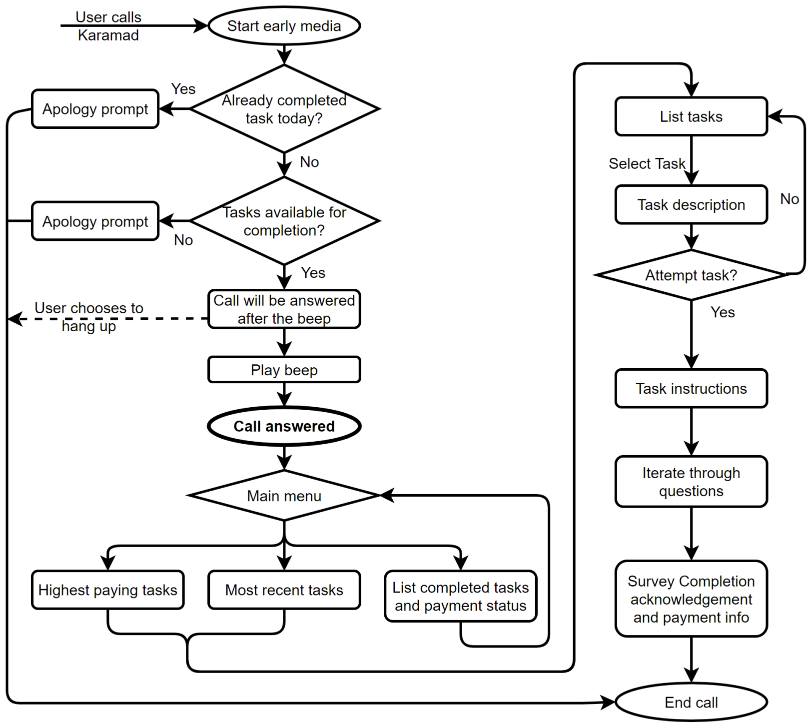

Figure 1: Simplified flowchart of worker IVR

the following adjustments to support the implementation of survey questions and efficient task navigation.

Prior work reports that multiple-digit input can be complex for users and suggests incorporating examples to address this issue [8]. We took the approach of restructuring multi-digit input prompts into multiple-choices by segmenting possible answers into ranges. For example, rather than asking for a user's age as multi-digit answer, we reframe the question into a multiple choice question with responses in the form of ranges such as <20, 21-30, 31-40, >40. This approach is less precise, but it is also less complex for users and less expensive overall as including usage examples in the question statement requires more airtime. Similarly, when MCQs exceed 9 choices, we divide into sets of 8 options followed by the "other" choice, leading to the next 8 options.

Karamad allows dynamic tasks where upcoming questions may depend on the responses to previous questions. Further, the visibility of tasks themselves to workers may also be conditioned on answers to prior tasks or questions answered. For example, an accessibility survey may only be made available to workers who completed the demographics survey and responded 'yes' to being visually impaired. The visibility of prompts and surveys are defined through the requester interface using a checkbox marking the questions as dependent and a drop-down menu to choose the dependency. Previous works [8, 49] also found a high incidence of users answering without listening to prompts. We made audio prompts unskippable in Karamad by disregarding any input while the prompts are being played.

3.4 Worker Interface

The IVR-based worker interface allows users to listen to the list of available tasks, and complete tasks of interest. As we do not employ missed calls, Karamad directly answers all calls. To throttle the traffic in the current deployment, we restricted each worker to attempt no more than one task each day. In addition, each worker can attempt a particular task only once. Figure 1 is a simplified flowchart of Karamad's worker interface. First, Karamad greets the worker during the early media stage and introduces the service (*"Greetings! Welcome to Karamad. Here you can earn mobile top-ups by completing posted tasks"*). It then uses the phone number of the caller to verify if they are eligible for attempting any available task. Users may be ineligible because: (1) they have already completed a task earlier that day, (2) they have already finished all tasks available on Karamad, or the available tasks have already gathered enough responses as required by the requester, or (3) the worker's profile does not match the dependency requirements of the available tasks.

When tasks are available for workers, they are informed that Karamad will answer their call after the beep and they will be charged for airtime from that point on. At this stage, workers can still choose to hang up if they do not wish to continue. If the worker is not eligible for attempting the available tasks, Karamad apologizes, and requests that they call back the next day (if they have already attempted a task that day) or call later (if they have attempted all tasks for which they are eligible). Karamad then politely hangs up without costing the worker any airtime.

After the beep, Karamad answers the call for the workers who decide to remain connected. Next, they are given the option to browse tasks or to listen to the payment status of their completed tasks. Tasks can be browsed by reward (highest reward first) or recency (most recently posted first). Based on their choice the list of task descriptions and reward values are played to them. When the tasks are sorted by reward, the amount is played before the task description and in the case of recency, the task description is played prior to reward. In the task description, workers are informed about the type of task, the kind of information they would be expected to provide, and response types involved in the task (e.g. MCQs, binary, audio recordings). When a worker selects a task, they are provided instructions for successfully completing it (e.g. only repeat the numerical digit mentioned in each prompt, record your responses in as quiet an environment as possible). Then the interface starts iterating through the questions/prompts until the task is finished. The worker is thanked and informed about their payment schedule.

3.5 Requester Interface

The requester interface is implemented as a web application modeled after MTurk's requester interface [1]. At a high level, requesters are able to create, edit, save, preview, delete, publish, and monitor tasks via the requester interface. As Karamad is a voice-based platform, there are two key differences from existing crowdsourcing interfaces. First, all instructions and prompts are specified via uploaded audio files. The text-based instructions for the tasks and individual prompts are solely for the convenience of the requester to keep track of the contents in each audio file. Second, because we wanted to ensure that workers are paid reasonably, the platform automatically calculates the minimum price of a task based on an estimate of the total expected time to listen to and respond to all of the questions/prompts.

4 EVALUATION

Our main research question was: "Can an unsubsidized IVR platform be used for general-purpose crowdsourcing in a manner that sustainably engages low-resource workers?". To answer this question, we evaluated Karamad along five main dimensions:

- **Viability:** Can the platform attract workers, engage them in lengthy calls as they complete tasks, and retain them for multiple interactions?
- **Task-support:** Can workers using the platform complete tasks of varying type, complexity and sensitivity?
- **Sustainability:** Are promised financial incentives and platform features (e.g. ring-back-tones) sufficient for workers to invest their time, energy, and airtime?
- **Reliability:** How timely, coherent, and consistent are worker responses?
- **User satisfaction:** Do workers value their interactions with Karamad as worthwhile, profitable, and fulfilling?

We used a mixed methods approach. We performed quantitative analysis of call logs for measuring service spread, user engagement, and retention with various task and question types. We assessed reliability based on the quality of worker inputs and pairwise comparison of responses to repeated and reordered questions. Finally, we conducted semi-structured telephone interviews to collect demographics information, user satisfaction, and other feedback.

All four authors are HCI4D researchers who have previously conducted a variety of surveys, data collection campaigns, and IVR deployments in Pakistan and elsewhere. On the basis of our own experience, existing literature on crowd work and IVRs, and based on the market value of surveys conducted on Mturk, we considered surveys conducted by the World Bank and the government of Pakistan to select a breadth of topics that would be useful to collect from hard-to-reach populations. We also came up with our own surveys on topics that we thought would have relevance to the demographics of typical IVR users. Other than surveys, we created 4 speech corpus collection tasks based on prior speech corpus creation efforts [31, 38, 48, 64].

In total, we designed 12 surveys, 4 speech data collection tasks, and 1 validation survey. Table 1 summarizes the tasks. The tasks were released to users in two batches. The first batch consisted of the first 8 tasks in Table 1. The main goals of the first batch were to introduce potential users to our system, collect basic demographic information to understand our workers, and test the usability of our system. After reviewing the results from our first batch of tasks, we developed the second batch of tasks that were longer, more challenging, and extended into potentially sensitive topics. For each task, two of the authors worked together to translate each survey to Urdu (if they were not already in Urdu)[2]. We paid special attention to ensure that the questions would be easily understood by our workers. We used the requester interface to create and deploy tasks on Karamad.

Institutional Review Board (IRB) approval was obtained before beginning our deployment of tasks. A disclaimer was added to Karamad to let the users know that their interactions are being recorded for research and quality improvement purposes. We limited the maximum number of unique responses for each survey to 300.

4.1 Tasks – Batch 1

The first batch of tasks was based on reports by the World Bank [4], and Pakistan Bureau of Statistics [42]. The initial surveys were limited in complexity, consisting of binary and multiple choice questions and fewer than 5 questions. Very few questions were open-ended (4 out of 71) and required spoken responses. We released the first batch incrementally to assess the attrition rate of our workers over time and to fix any remaining bugs. Longer surveys with questions having as many as 15 choices and disability-related questions were released three weeks later. We also deliberately included surveys with sensitive questions about participant's habits and knowledge of potentially taboo topics (e.g. sexually transmitted diseases) to figure out whether our users were comfortable with slightly longer surveys and more sensitive questions. Finally, we limited the eligibility of the employment surveys to users who responded as being employed and literacy surveys to users having less than a high school education, in the basic demographics survey.

4.2 Tasks – Batch 2

After analyzing the results from the first batch of tasks, we designed two more sets of tasks and a validation task. The first set

[2]Pakistan is home to 74 native languages. Urdu is the national language and over 109 million (53%) people in Pakistan speak Urdu [16].

Table 1: Survey characteristics. The first 8 surveys were released in batch 1 and the remainder in batch 2.

Survey	Questions	Dynamic Paths	Questions in Paths	Duration (minutes)	Est. completion cost (PKR)	Reward (PKR)	Eligible	Attempted	Completed
Basic demographics	5	1	5	4	16	50	300	312	300
Accessibility	8	4	5-8	4.4-6.5	20-28	50-80	300	308	300
Health	17	6	3-17	2.5-9.3	12-40	40-170	300	316	300
Disability	13	12	7-13	6-10.2	24-44	70-130	300	314	300
Ethnicity/Language	8	6	5-8	6.3-8.6	28-36	50-80	300	305	300
Employment	7	1	7	5.92	24	70	120	93	90
Literacy	7	4	5-7	3.9-5.1	16-24	50-70	211	177	176
Formal Education	6	3	2-4	2-4.8	8-20	40	300	303	300
HIV/AIDS	17	3	1-17	1.6-17	8-68	40-170	300	175	174
NCD Risk Factors (Eating/Drinking)	16	16	9-16	10.1-18	44-72	90-160	300	218	212
NCD Risk Factors (Smoking/Activity)	15	32	7-15	8.2-18.7	36-76	70-150	212	178	175
Blind General	19	5	1-14	1.7-11	8-44	40-140	300	227	216
Blind Work	14	4	7-10	5.6-7.6	24-32	70-100	300	166	164
Blind Phone	22	24	12-22	10.4-18.5	44-76	120-220	300	209	201
Isolated Digit Corpus Counting Ascending	20	1	20	17.3	72	200	300	174	152
Isolated Digit Corpus Counting Descending	20	1	20	17.3	72	200	152	126	123
Isolated Digit Corpus Counting Random	20	1	20	17.3	72	200	123	103	100
Speech Corpus	13	1	13	12.6	52	130	300	126	106
Validation	20	1	20	16.8	68	200	300	274	250

of tasks focused on sensitive but important topics that are typically embarrassing or taboo. Because prior work in low income countries has shown that IVR-based surveys have less social desirability bias as compared to telephone surveys by interviewers, respondents are more likely to answer truthfully regarding sensitive topics [17, 44, 56]. We designed surveys around HIV/AIDS and non-communicable diseases (NCD). The HIV/AIDS survey is based on the World Bank health report [4] and in it we asked questions related to knowledge of HIV prevention methods, misconceptions regarding it, and discriminatory attitudes towards people living with HIV. The NCD survey adapts questions from the World Health Organization Stepwise Approach to Surveillance of NCD risk factors (STEPs) surveys, which has previously been successfully conducted using IVRs in multiple low- and middle-income countries of Bangladesh, Uganda, and Tanzania [20, 44, 60]. Because many of our workers are blind, we also created three surveys about living with blindness and its challenges.

The second set of tasks was focused on collecting Urdu speech corpora. We designed tasks that collect isolated digits and complex sentence-translations in free speech form. Though many approaches could be taken to collect speech data, for isolated digits, we designed three tasks based on prior works collecting corpora of isolated Urdu digits [38, 48] that asked participants to record spoken digits in ascending, descending, and random order. We also asked participants to repeat the prompted sentences in their mother tongue, which could be different from the language in the prompts.

Our final task was a survey intended to test the internal validity of participant responses. The concern being that workers may be inattentively pressing buttons or not making audio recordings.

4.3 Task Pricing

We priced our tasks so that participants earn roughly 1.5 to 2x the maximum amount of phone credit cost that we estimate a worker spends on airtime while completing each task. Our pricing is not precise due to non-linear pricing schemes, hidden fees, payment portal restrictions, and a general effort to remain conservative in our estimates of airtime costs. To estimate the per minute airtime cost, we manually explored all of the mobile networks in Pakistan and found the most expensive network cost and used that as the user's cost basis. We found that the Jazz network has a package that charges 3.20 PKR/minute, plus a government sales tax of 19.95%, plus a call setup fee of PKR 0.15 for the first minute [2]. Since our tasks have conditional questions and different paths of questions based on prior answers, each task has a minimum to maximum number of prompts (indicated by the "Questions in paths" column in Table 1). We calculate the expected time to complete a task based on the average time taken to listen and respond to prompts across all question paths.

Based on our formulation and considerations, Table 1 summarizes the estimated cost of each task and the completion reward. The pricing of rewards for tasks we requested on Karamad were set generously for several reasons. First, we wanted to ensure that our

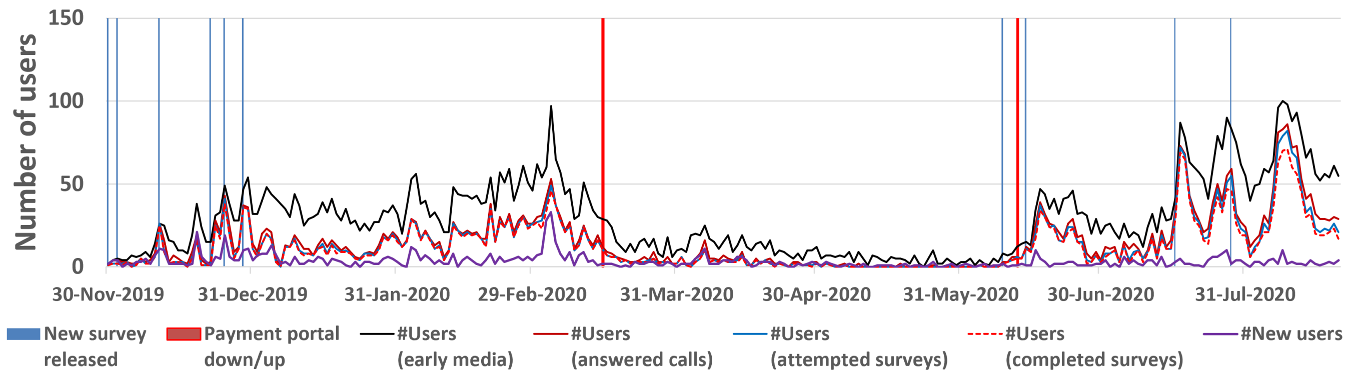

Figure 2: User traffic of Karamad deployment

workers were not losing money due to airtime costs, particularly if users made mistakes while learning to use the system. Our workers being very poor have low tolerance to risk. Second, we wanted to ensure that they were paid at least minimum wage [59]. Third, we wanted to pay at a comparable market rate per task relative to existing tasks on MTurk [24]. Given the minimum wage in Pakistan for unskilled labor is set to PKR 15,000 per month, for 8 hours a day, and 26 days a month, the per-hour wage is roughly PKR 72 (0.43 USD) per hour. Based on these considerations, we paid roughly PKR 1,400 (8.43 USD) per hour, which is 20x federal minimum wage in Pakistan. This is a modest sum even in the context of Pakistan because the minimum wage is very low. To put this in perspective, the cost of the smallest monthly internet package for Ufone, a local mobile network provider, is 390 rupees or approximately 5 times the minimum wage. In absolute terms, workers spent on average 81 minutes using Karamad in total during the course of our deployment. Our pricing is also roughly in line with recent closely related IVR-based pilot studies conducted in India (we paid approximately $0.042 per question and approximately $8.43 per hour after accounting for airtime costs compared to $0.046 per question for Learn2Earn [61] and $9.81 per hour in the case of ReCall [63]).

4.4 Deployment

Karamad was launched on Nov 30, 2019 and remained active until Mar 15, 2020, when we took it down due to problems with our payment portal. It was made live again on Jun 11, 2020 until we finally concluded the initial deployment on Aug 20, 2020. Overall, it remained active for 176 days (nearly 6 months). During this period we posted 19 tasks on Karamad, of which 14 were independent while 5 were only available to a subset of workers based on their responses to questions in previously completed tasks. During the initial weeks of deployment we found that the majority of tasks were being completed by the same handful of individuals. To encourage diversity and more opportunities for new users, we imposed a limit of one task per day per user. This limit was imposed on Dec 24, 2019 and stayed active till the end of deployment.

We never advertised Karamad or formally recruited any workers. We deployed Karamad by making it available on a phone number where a popular hotline, Sawaal [50], was hosted previously that had been offline for 4 years. Within a day, we started receiving phone calls. We probed this in user-experience interviews (section 4.5.7) and found out that some users of the previous service (most of them blind), kept calling the number to see whether the old service was active or not. As soon as they found out that a new service is live they spread the news through their offline networks to their peers, most of whom were also visually impaired. The payment mechanisms employed in this deployment restricted us to 20 transfers per day (5 for each mobile network, totaling not more than PKR 10,000 each day). Therefore, we combined payments into batches and did not incorporate features that are associated with viral spread and exponential growth like sharing and forwarding [49–51] or referrals [62]. Also, our deployment did not subsidize airtime for the users and did not employ toll-free numbers or missed calls [49–51, 62].

4.5 Results

Despite the lack of any advertisement, airtime subsidy and viral interface features, users discovered Karamad on their own and spread the word about it to others. Our interview participants informed us that their peers had also posted about Karamad on other voice forums. Over the 6 months of deployment Karamad reached 958 unique users out of whom 725 attempted at least one task and 671 completed at least one task. These users called Karamad 16,355 times and contributed 3,939 survey responses (on average 213 responses for each of the 19 tasks). Through these responses we gathered 43,006 answers to 267 questions, including 16.20 hours of speech data in response to open-ended questions, and 1,360 audio translations via translation tasks.

4.5.1 Demographics. The first survey posted on Karamad asked the users to report their demographics. Based on 313 responses from participants who self-selected the demographics survey we found that 89% of our respondents were male, and 11% were female. Among them 45% reported their ages to be between 20 and 30 years, 16% between 30 and 40 years, 9% above 40 years, and 3% less than 20 years. Also, 61% of our users were unemployed. In terms of education, we had 40% non-literate users, 15% with up to 5 years, 14% with up to 8 years, 17% with up to 10 years, while 14% with more than 10 years of education. 85% of our users reported themselves as being visually impaired.

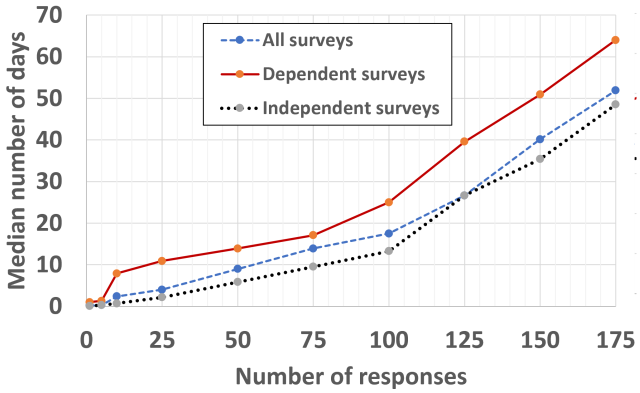

Figure 3: Median time to gather responses

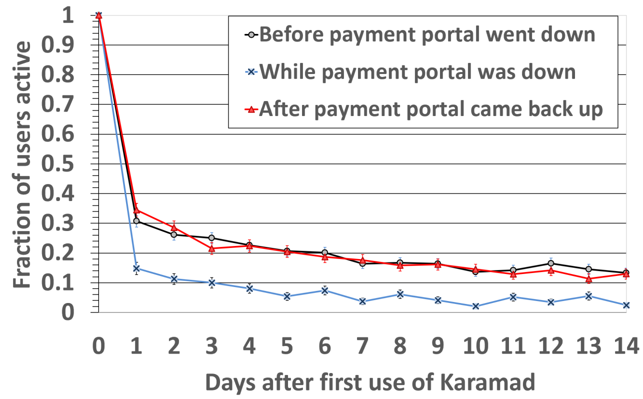

Figure 4: User Retention

4.5.2 Spread. We investigated the spread of the service further when we conducted manual telephone interviews with our users. We found out that the people in the blind community were informed about Karamad by one or more of their blind peers. These peers were described to us as prolific users of voice services who keep calling the phone numbers associated with various voice forums and inform others as soon as they find useful services.

In Figure 2 we observe that close to 50 users on average called Karamad every day (solid black) and new users (solid magenta) were being added to the service throughout the 6 months of active deployment and even during the phase when the service was down due to the unavailability of our payment portal (between the thick red vertical lines). We see spikes in the number of daily users corresponding to the dates of launch of surveys (thin blue vertical lines). This is indicative of offline communication, as we did not advertise the availability of new surveys on the platform. We also observe that users continued calling Karamad during the months of downtime although most of them hung up (or were disconnected by Karamad) during early media before their call was answered.

A key observation from Figure 2 is that the number of users who reached the early media stage on any given day is significantly higher than those who reached the answered stage, where they were charged for their calls. In fact, 48% of all users who reached the early media stage, were disconnected before their calls were answered (mostly) due to the unavailability of tasks for which they were eligible. Moreover, of the users who *were* eligible for tasks, 87% of them were able to attempt a task (solid blue) and 93% of those who attempted a task were able to complete it (dotted red). This shows that early media was able to offload a large fraction of calls that would have wasted user airtime while enabling users to efficiently complete tasks. This trend also holds for the new users who joined the service during our deployment period.

4.5.3 User Engagement. We find that the 777 workers whose calls were answered by Karamad, spent 24,926 minutes (avg. 32.0 minutes/worker) interacting with the service in these calls. Overall, workers spent PKR 80,000 (USD 533) out-of-pocket to interact with the service or approximately 0.68 USD per worker. Karamad informs each user at the completion of the survey that they will be paid within a few days after their survey has been reviewed by the requester. Each user essentially worked on credit until they got paid. The 681 workers who completed at least one task spent 16,834 minutes (avg. 24.7 minutes/worker) completing their tasks.

Of the 16,440 calls to Karamad, 11,451 (69.6%) were handled and disconnected in the early media stage without costing the user any airtime. Of the 11,451 disconnected calls, 4,603 (40.2%) were disconnected due to unavailability of tasks for the workers who called, 5,217 (45.6%) were disconnected as they had already completed their quota of one task for that day, and 1,631 (14.2%) calls were hung up by the callers. Out of 16,440 calls, only 4,989 (30.3%) calls progressed to the answered stage and in 3,988 (79.9%) of answered calls, the workers attempted tasks. Karamad was able to handle most of the unnecessary call traffic during the early media stage, without charging the workers any airtime expenses. This also reduced the call load on the system.

4.5.4 Income. Table 3 shows user income statistics on Karamad. Workers earned PKR 637.4 on average (median income = PKR 350). The difference between the median and mean is due to the skew where 186 (out of 681) workers just completed one task, while 154 completed more than half (10) of the available 19 tasks. On average, each worker contributed 6 tasks (median = 4 tasks). The net income per user, after accounting for the amount that they spent on airtime, comes to an average of PKR 490 per user (median = PKR 266). However, for the top 10% workers in terms of time spent completing surveys, each worker earned on average PKR 1,347 (median = PKR 1,350) after spending 142 minutes on average (median = 135 minutes). Therefore, the total amount earned by the 681 workers from Karamad comes to PKR 328,094 (PKR 90,262 for the top 10%). These users spent PKR 98,976 worth of airtime (PKR 38,068 for the top 10%) completing tasks on Karamad. It is noteworthy that users were willing to invest this time, money, and effort completing each survey several hours (sometimes days) before receiving their payments.

4.5.5 Response time. The time required to gather n responses to the tasks varied across tasks as we did not have any mechanism to advertise the availability of new tasks to the users. The time also varied between independent and dependent tasks because the number of eligible workers is much smaller for dependent tasks. Figure 3 shows the median time to gather the first n responses

Table 2: Usage statistics

Statistic	Count	Users
Total calls received	16,440	902
Calls in which user disconnected in early media stage	1,631	479
Calls in which system disconnected due to unavailability of tasks	4,603	304
Calls in which system disconnected as user had reached the daily quota of tasks	5,217	435
Calls that were answered	4,989	777
Total time spent by users in all answered calls	24,926 minutes	777
Calls in which a task was attempted	3,988	725
Total time spent by users in calls where the attempted tasks	24,926 minutes	777
Calls in which a task was completed	3,399	681
Total time spent by users while completing tasks	16,834 minutes	681
Total tasks attempted	4,123	725
Total tasks completed	3,939	681
Total questions answered	43,006	719
Total money paid to users (approximately)	$2,836.85	595

to the 19 tasks. Overall, the median time to get the first response was 5 hours overall (2.5 hours for independent and 24 hours for dependent tasks). Similarly, it took 7 hours overall (independent: 3.5 hours, dependent: 31 hours) to gather 5 responses and 57 hours overall (independent: 16 hours, dependent: 190 hours) to gather the initial 10 responses. Beyond the 10 responses mark, the completion time grew nearly linearly reaching 60 days to gather 175 responses overall (independent: 49 days, dependent: 64 days). We believe that the initial 10 responses require an unusually low amount of time as these are contributed by a group of well-connected, prolific users. Some of this could also be due to the users who have multiple phone numbers registered against their name. Adding explicit advertisement (e.g. by calling up the workers or sending them text messages), has the potential to smooth out the abrupt transition around 10 responses and also to decrease the gradient of the completion time for gathering responses.

4.5.6 User Retention. Figure 4 shows the user retention graphs of Karamad. Each graph displays the fraction of users who continued interacting with Karamad k days after their first interaction (where k=1 to 14) out of the cohort of users who started using Karamad at least k days before the last date of deployment. This cohort represents users for whom we have observed a window of at least k days when they could have called Karamad. For example, a user who starts using Karamad two days (48 hours) before the end of the deployment only has a chance to use it for two days. The graphs align users by their kth day of interaction with Karamad. E.g. k=0 is the day when users call Karamad for the first time (all users interact with the service on their first day). k=1 shows the fraction of users who call Karamad (at least once), one day after their initial interaction with the service. For each k, we model the cohort of n_k users who could have called Karamad as n_k, i.i.d, and unweighted Bernoulli trials. The error bars show the standard error of the sample mean of the binomial distribution. Because of the large sample sizes, the bars are barely noticeable.

We divided the deployment into three phases, the period before the payment portal was down due to technical difficulties, the period in which the payment portal was unavailable and new tasks were not posted, and the period after the service was restored until the end of deployment.

User retention during the period when the payment portal was down, is the lowest, but non-zero. Between 2% to 5% of all users kept calling back during this period, presumably to check whether the service has been restored or not. User retention during the two active periods follows very similar trends. Between 30% to 34% users return to the service a day after their initial interaction, 27% return on the second day, and after that the drop is nearly linear for the next 12 days. Two weeks after the initial interaction, 13% of all users keep calling Karamad. Compared to other voice forums, Baang [49] – a community platform and social network – retains nearly 17% of its user-base after two weeks while Polly [51] – a peer-to-peer viral entertainment service – retains 4%. The user retention of Karamad, is similar to the retention exhibited by social platforms. We attribute this high retention to the financial incentives of Karamad, which shows that financial incentives, like social ones, have the ability to retain the user-base of voice-forums.

4.5.7 Telephone Interviews. We contacted 140 users who had performed tasks on Karamad to conduct semi-structured telephone interviews. 35 users agreed to be interviewed. All calls were made using Karamad’s caller ID. The interviews were recorded with the permission of the respondents. One author also took notes of the conversation. The interview included questions about demographics, accessibility, feedback about the service and reasons for errors in user input (discussed in the next section).

Demographics: All 35 respondents identified themselves as being visually impaired with 26 of them being completely blind. Only 2 female participants agreed to be interviewed. The respondents belonged to 21 different districts of Pakistan across 3 different provinces (Punjab, Sindh and KPK). In terms of educational background, 6 were non-literate, and 5 had only received basic religious education. The remaining 24 had received formal education. Of these, 3 had up to 5 years, 15 had up to 10 years, and 6 had more than 10 years of education. In terms of employment, 20 respondents were unemployed, of which 2 were students. Others reported various professions including teachers at religious schools and schools

Table 3: Income and time

	# Surveys completed	Time spent (minutes)	Earned from Karamad (PKR)	Spent on airtime (PKR)	Net earned (PKR)
Average	6	36.9	637.4	147.7	489.7
Median	4	17.0	350	68	266
Average (top10%)	15.6	142	1,915.4	568.2	1,347.2
Median (top 10%)	17	135	1,950	540	1,350
Total	4,027	24,744	427,070	98,976	328,094
Total (top 10%)	1,048	9,517	128,330	38,068	90,262

for the blind (6), imam of a mosque (1), salesman (1), voice artist (1), daily wage workers (3), and small home-business owners (3) that they conducted with the help of family. In terms of monthly income, 20 reported having no income and being financially dependent on their families and welfare trusts, 2 reported less than PKR 3,500, 5 between PKR 3,500 – 10,000, 6 between PKR 11,000 – 30,000, and 2 daily wage workers with fluctuating income. Of our 35 respondents, 21 mentioned having access to internet even though only 12 had access to smartphones; the remaining respondents had mediated access to the internet with the help of their family. Most respondents (23) owned feature phones, 10 had access to smart phones, 2 owned both smart and feature phones. Only 6 respondents had assistive software on their phones.

Most visually impaired respondents appeared quite protective of their community. They believed Karamad to be a service specifically for blind people and suspected that some non-blind people may be taking advantage of the service by pretending to be blind. They gave us advice regarding creating checks to filter out such users. One respondent said:

> *"A blind person usually likes to talk a lot and is very excited when talking. A person who has vision is mostly to the point. So, you can just listen to the recordings of your users and block those who are giving succinct responses. As you made this service for the visually impaired who cannot get normal jobs. You should at least have four recording questions to judge from the responses if a person is blind or not"* (Male, 29, Micro Entrepreneur, 10 years of education)

Other respondents (10) reported hearing about the service on a voice forum where a user of Karamad had created a post about it. He had also forwarded this recording to his family and contacts. The two female respondents mentioned that although they never used any voice forum before, their brothers (who are also visually impaired) use such services and they told them about Karamad.

Feedback on Karamad: We asked the respondents for their feedback about Karamad, getting paid for completing tasks, and the kinds of tasks posted on the platform. Overall, we found that Karamad provided workers with a sense of independence and usefulness. Users expressed the desire for more tasks and task diversity and provided considered feedback regarding platform features.

A large fraction (20 out of 35) of respondents, especially blind users, explained that they were dependent on their families and welfare organizations for financial support, and had experienced personal income for the first time through Karamad. One respondent said:

> *"It was very beneficial for me because I am financially dependent on my brothers and family. Because of this mobile top-up I now don't have to ask them anymore."* (Female, 38, small shop owner, 10 years of education)

More than half (20) of our respondents said that the surveys are very good and help them improve their knowledge. Some respondents also criticized the questions as being too easy. A few respondents (2) asked us to include more surveys on general knowledge. Respondents also asked us to improve the diversity of survey topics. One respondent explained:

> *"There should be some Islamic questions or content especially on Fridays."* (Male, 28, Quran teacher, 5 years of education)

Respondents emphasized that new tasks should be added more frequently, and that such schedule should be announced to them along with the number of questions in the upcoming tasks so that they can maintain mobile balance for their Karamad calls. He explained:

> *"You should tell before the tasks how many questions you are going to ask in next task so that we can maintain call balance."* (Male, 44, unemployed, 10 years of education)

A respondent complained that they have to keep checking for new surveys by calling Karamad. Respondents asked us to add notifications for new tasks either via SMS or missed calls from the hotline numbers. Another respondent asked us to come-up with a mechanism to save progress of the tasks in case the calls gets disconnected due to low balance or issues with mobile networks.

Mobile top-up mechanism: Respondents also gave us detailed feedback on our payment mechanism and timeline. Most respondents (31 out of 35) said that the income that they make through Karamad means a lot to them since they do not have to ask anyone for mobile top-ups anymore. They have enough balance now to subscribe call and internet packages. We did not receive any complaints about delays in payments.

We also asked respondents how they spent their mobile top-up and 33 respondents mentioned using top-up earned from Karamad on subscribing to call and internet packages. While 2 respondents stated using some by themselves and selling the rest to their peers in exchange for money. One respondent stated:

> *"I used some by myself and then sold the rest to other people. The balance that I sold to others was around PKR 700 – 800"* (Male, 27, unemployed, non-literate)

We also encountered instances where users had been calling Karamad using multiple phone numbers. A few respondents complained that some of their peers have been taking advantage of this service by working from their all SIM cards. One respondent even gave us a suggestion on how to fix this:

> *"You should have some check and balance so that same users do not access the service from different numbers. You should register users on this service against their identity card number so that you can send money in their mobile wallet accounts. Because you can only make one wallet account against an identity card number while you can own up to five SIM cards against one identity card number"* (Male, 23, unemployed, non-literate)

Nearly half (14 out of 35) of our respondents asked the money to be sent to mobile wallet accounts rather than as top-ups because they could use this money to buy things. They stated that current top-up amount is more than their monthly usage and mobile top-up expires after some time. One respondent further explained:

> *"Money should be sent to the mobile wallets because often it was too much for my usage and it expires after 90 days."* (Male, 26, daily wage worker, more than 10 years of education)

4.5.8 Quality and Reliability of Responses. Quality control mechanisms are an important aspect of crowdsourcing platforms, particularly given the perception of potential unreliability of workers from low-resource communities. We created a validation survey that seeks to determine the consistency of responses provided by the workers. The validation survey consisted of 4 MCQ's, 4 Yes-No, and 2 open-ended short answer questions (Table 4). Each of these 10 questions either had their answers shuffled or questions rephrased to create another 10 questions. We gathered a total of 250 worker-responses for the validation survey.

We find that 30.4% of all workers answered all 10 pairs of questions with complete consistency, 27.2% had one inconsistent pair, 18.4% had two, 5.6% had three, 14.8% had four, while the remaining 3.6% had more than 4 inconsistent pairs of responses. Therefore, 96.4% of all workers responded with no more than four inconsistent answers. We found that the workers tended to be more consistent with their responses to binary questions (74.4% were consistent, 20.8% had one inconsistency, 4.0% had two, and 0.8% had three) compared to multiple choice questions (39.2% were consistent, 26.0% had one inconsistency, 16.0% had two, 4.0% had three, while 14.8% gave inconsistent answers to all MCQs). The lowered consistency given more options may reflect the increased cognitive load required in text-free navigation tasks as reported previously by Medhi et al. [37]. Among the multiple choice questions the question about internet usage received the highest fraction of inconsistent responses, followed by the questions about education, ethnicity and marital status. We discovered some of the reasons for the higher inconsistency among MCQs via our telephone interviews as we discuss later in this section.

The open-ended audio questions were about the preferred language and location of the workers. The recorded audio varied between 5 to 20 seconds. One of the authors listened to all 1,000 files, transcribed the contents in the recording, and then assigned them tags. Each file was tagged as one of "non-content" if it did not contain actual spoken content (e.g. it contains silence or noise), "relevant" if the worker provided a response that is relevant to the question, "irrelevant" if the worker's response was not related to the question. After each individual file was coded as above, meta-tags were assigned to each pair of questions based on whether both answers matched or not. A "match" tag was assigned if both responses matched, and "no match" was assigned if answers did not match. We also had a "partial match" tag for the situations where some of the provided information matched however there was additional information in one of the responses which did not match (e.g. mentioning the preferred language in one response to be "Urdu" and in the other as "Urdu and Punjabi"). We also had tags to indicate that one or both files have either no content or irrelevant content. In our location question we had asked the workers to provide us with the name of their city or district. This was done to handle situations where a worker does not know the name of one or the other. However, this resulted in an apparent inconsistency, which we accounted for by assigning a "hierarchical match" tag.

For the first occurrence of each question (original question), out of total 500 files, 432 (86.4%) were tagged as relevant, 61 (12.2%) as non-content, and 7 (1.4%) as irrelevant. Similarly, in the second (rephrased) occurrence of each of the 500 questions, 440 (88%) were tagged as relevant, 54 (10.8%) as non-content, while 6 (1.2%) as content irrelevant.

In terms of meta-tags, for the 250 pairs of location questions, 169 (67.6%) had a complete match, 8 (3.2%) had a partial match, 16 (6.4%) had a hierarchical match, 17 (6.8%) did not match, while the remaining 40 (19%) had no content or irrelevant content in one or both files. Therefore, the overall consistency between the answers was 193 (91.9%) out of 210 valid pairs, while 17 (8%) valid pairs mismatched. For the preferred language question, out of the 206 pairs that had valid content in both files, 150 (0.60%) pairs had a complete match, 23 (9.2%) had a partial match, while 33 (13.2%) of the pairs had a mismatch. Therefore, the overall consistency between valid pairs was 83%. We asked questions in our interviews to understand the reasons for low consistency between responses to MCQ questions. Of our 35 surveyed workers, 15 had mismatches in their answers to the internet usage question, while 8 had mismatches in their marital status question. Most users attributed their inconsistencies to confusion or errors. All users, whose marital status mismatched, mentioned that they mistakenly pressed the wrong button while answering the question.

For the internet usage mismatch, while most users reported mistakenly pressing the wrong key, some mentioned that they were confused because they have recently changed their internet enabled devices and their internet usage patterns are in flux. One user explained:

> *"I used to own a smartphone and had access to the internet but now I have a feature phone so I can't access the internet anymore. That is why my answers did not match"* (Female, 38, micro-entrepreneur, 10 years of education)

Another respondent explained that he was confused with this question as he was not sure what would count as "internet usage":

Table 4: Questions in the validation survey

Question type	Topic	Answer choices
Binary	Have you ever been surveyed before?	Yes/No
Binary	Can you speak and understand English?	Yes/No
Binary	Is there any language that you can read?	Yes/No
Binary	Are you seeking paid work?	Yes/No
MCQ	Education	(1) Currently attending school, (2) Attended school in the past, (3) Never attended school
MCQ	Recent internet usage	(1) Almost every day, (2) At least once a week, (3) Less than once a week, (4) Not at all
MCQ	Marital status	(1) Widowed, (2) Divorced, (3) Unmarried, (4) Married
MCQ	Ethnicity	(1) Saraiki, (2) Balochi, (3) Muhajir, (4) Sindhi, (5) Pathan, (6) Punjabi, (7) Other
Short Answer	Location (city/district)	Open-ended
Short Answer	Preferred language	Open-ended

> *"I was confused about how to respond. As my phone is Symbian-based, I can only perform limited tasks on the internet. I can only download audio books from a website and I only use the internet for this purpose. Also, I cannot use different social media platforms because my phone does not support it"* (Male, 44, unemployed, 10 years of education)

5 DISCUSSION

The primary motivation behind our work was to explore the challenges and opportunities of making general crowd work available to and accessible for underserved populations in a sustainable manner. In this section, we reflect on the results of our work, its limitations, and the research opportunities in crowdsourcing.

Over the last decade, IVR platforms for underserved communities seem to have settled on several de facto standards of interface design. These include airtime subsidies, social features, and advertisements to attract and retain users. Prior work shows that lack of airtime subsidy leads to low call volume and engagement [51, 62], which causes IVR services to struggle with long-term sustainability [63]. Successful IVR forums have also engaged and retained users through social features (e.g. voting and comments) [49, 62], spreading mechanisms such as the ability to forward content to friends [49, 50], incentivized referrals [61], and advertisements [41]. Contrary to these prevailing norms, we designed and deployed Karamad without subsidies, social features, spreading mechanisms, or advertisements. Our work advances the existing HCI literature on IVR services for underserved populations by showing that in the absence of such features, paid crowd work can similarly achieve trust, engagement, spread, retention, and self-training.

Beyond the questions of accessibility and sustainability, we have observed that crowd work opportunities are welcomed by our users, that they can be easily assimilated into daily routines, and that the platform and associated income is viewed positively by low-literate, low-income, and blind populations. Our users enthusiastically engage with Karamad, actively spread the word in their communities, and invest significant up front effort, time, and money. Our surveys also reveal that workers welcome the intellectual challenges of answering questions as well as the resultant income that they value as an enabler of financial independence and opportunities.

5.1 Compensation and Fairness

In our design and deployment of Karamad our foremost consideration after usability was how much to pay workers to complete tasks while protecting them in our study. We carefully considered the nuances described in prior HCI works that explored the asymmetries in labor relations, the ethics of perpetuating inequality, and both of these issues in the context of globalized markets [14, 26, 53]. Our eventual decision to pay 20 times the Pakistan federal minimum wage was largely predicated on the ethical grounds, rather than an attempt to minimize price. We intentionally 'overcompensated' our workers, but our pricing is generally in line with similar IVR-based pilot studies recently conducted in India [61, 63].

We do not believe that such a high amount is *required* for robust participation from workers in long term operation due to increased familiarity and reduced risks. On the contrary, our survey findings echo Ekbia and Nardi's observation [15] that, beyond purely financial reasons, there are other motivations such as feelings of empowerment and independence that promote participation in heteromated labor arrangements. Based on prior work and given the overall enthusiasm that our workers expressed, we speculate that as long as airtime is subsidized very little financial compensation would be needed to motivate them. However, although lower prices would enable local employers to participate in our system, lower prices could also contribute to exploitation by more powerful international employers as previously described regarding crowd work [15, 27, 53].

We plan to further investigate the sensitivity of crowd workers to task prices, but the more critical question is how to balance between fairness and economic efficiency given the asymmetries in labor relations. The dilemma of how designers should think about crowd work in relation to political economy is a broad issue that some HCI researchers have begun to explore [26, 39, 58]. Within the scope of our work, we believe that positive emotional responses toward Karamad are a good outcome, but they should not be used as a palliative replacement for compensating workers fairly. Therefore, while we will explore more inclusive and transparent mechanisms to facilitate participation and voice [6], we will continue paying workers more than the required local minimum wage and support local employers by subsidizing their fees with those from international employers.

5.2 Quality of Crowd Work

We found high consistency between worker-responses for binary questions (74.4% workers responded with complete consistency, 20.8% had one inconsistent response, 4% had two and 0.8% had three) and for open-ended audio questions (91.9% and 83%). For MCQs the consistency was reduced (39.2% workers responded with complete consistency, 26% had one mistake, while 16%, 4% and 14.8% had two, three and four inconsistent responses). Our surveys indicated that the inconsistencies in MCQ responses were largely due to confusion about questions and erroneous key presses. Therefore, we do not believe that low task quality is a fundamental limitation of our platform or workers.

Beyond improving how questions are phrased and simplifying prompts other strategies can be adopted to improve the validity and consistency of crowd work on Karamad. Karamad did not include any form of worker-training in contrast to existing platforms like Samasource [54] that employ training to improve the the quality crowd work. Assigning ratings and providing feedback to workers can also improve task quality. We find a strong correlation between the fraction of inconsistent, empty and irrelevant audio content and inconsistent responses to the quantitative questions (binary and MCQs) across individuals. For example, 47% of all audio pairs contributed by workers who had contributed 50% or more inconsistent quantitative pairs, ended up containing empty or irrelevant recordings. This fraction was only 2.6% for workers who had not contributed any inconsistent responses to quantitative pairs of questions. Given these observations, we believe that assigning a rating to each worker based on their responses to validation surveys and based on the employer-feedback can improve the quality of crowd work. Ratings can also be tied to the hourly rate of individuals following the model of popular crowd platforms. Other strategies to improve task quality include inter-worker-agreements, and cross-validation and majority voting. Users of IVR services are also known to improve in their efficiency and capacities for more complex tasks as they grow more familiar with the platform [67].

5.3 Unique Challenges and Opportunities

Nearly three months of Karamad's deployment overlapped with the surge of COVID-19 in Pakistan. During this time, we were exposed to novel use cases for our ongoing work during the COVID lockdown. The lockdown had serious consequences for small business owners, daily wage workers, and skilled labor. The very structures supporting their daily income were paused due to the lockdown. Under such circumstances, remote mobile phone-based crowd work presented an opportunity for reaching such populations and providing them the means for earning supplemental income.

Additionally, research that is dependent on surveys and face-to-face interviews with remote populations was placed on hold because it was nearly impossible for teams to travel. During the lockdown, we were aware of several research groups and survey firms who resorted to manual telephone interviews. While such interviews are certainly a fallback in the absence of other options, anecdotally, the response rate is generally very low. In our work, only 35 out of 145 contacted users agreed for a live interview with a human interviewer. Despite engagement with Karamad, our users (especially the female users) were unwilling to talk with a human interviewer. We believe that an interaction with a machine provides a sense of security, anonymity, and convenience that our target users do not feel when conversing with a person.

Building on prior work on crowdsourcing behavioral research [19, 30, 36, 47], we found high response rate and survey completion rate of Karamad's surveys on sensitive topics including hepatitis and tuberculosis (attempted by 316 workers, completed by 300), and HIV (attempted by 175 workers, completed by 174). In our interviews, respondents did not mention any discomfort while attempting surveys on these potentially sensitive topics. We believe that voice-based crowdsourcing provides a unique opportunity, beyond accessibility, for engaging difficult to reach respondents at scale regarding sensitive and taboo topics that they would not discuss with a person directly, or over phone.

5.4 Study Limitations and Future Directions

While we developed and deployed both the worker and requester interfaces for Karamad, we only tested the worker interface with actual users. The requester interface was employed by our team to post tasks on Karamad and was improved in the process based on internal feedback. Based on our experiences and user feedback, our platform could be improved in several ways. To keep the interface simple for users, we used phone numbers as a proxy for their identity. This currently lets users call the platform from multiple phone numbers, which could significantly skew survey data. This issue can be mitigated by employing a one-time registration and requiring the user to enter their 13-digit national identity card number. Future deployments of Karamad will support a resume-task feature that will help users when they are disconnected due to low balance or poor coverage. Similarly, SMS or missed call-based notifications will be sent to users as new tasks become available to prevent them from having to repeatedly call Karamad and still give them a chance to attempt the limited number of first-come-first-serve tasks.

6 CONCLUSION

In this paper, we explored the potential of voice-based interaction as a means for including marginalized populations in crowdsourcing. We explored the design of, Karamad, an IVR-based crowdsourcing platform that supports a variety of useful data collection tasks including surveys and speech corpus collection, and is especially accessible to underserved and hard to reach populations. In a 6-month deployment of Karamad, we found that, without any airtime subsidies, training, or advertising Karamad's financial incentives were sufficient to attract 725 unique users by word of mouth who completed 3,939 tasks and earned an average of PKR 637 per user. Karamad is currently live for beta access, please contact the authors for more information.

ACKNOWLEDGMENTS

We thank our reviewers for their insightful feedback that improved our paper, and our research assistants and associates Bilal Saleem, Abdullah Sabri, and Zain Tariq for spending lengthy hours translating voice prompts, annotating data, maintaining the system, and providing feedback on the interface.

REFERENCES

[1] Amazon. 2020. https://www.mturk.com/.

[2] Pakistan Telecommunication Authority. 2020. https://www.pta.gov.pk/en/media-center/single-media/deduction-of-charges--taxes-on-prepaid-re-load-recharge-and-tariffs-170120.

[3] Meezan Bank. 2020. https://ebanking.meezanbank.com.

[4] World Bank. 2019. World Bank Health Report. https://microdata.worldbank.org/index.php/catalog/3411/related-materials.

[5] World Bank. 2020. World Bank Poverty. https://www.worldbank.org/en/topic/poverty/overview.

[6] Chris Bopp and Amy Voida. 2020. Voices of the Social Sector: A Systematic Review of Stakeholder Voice in HCI Research with Nonprofit Organizations. *ACM Trans. Comput.-Hum. Interact.* 27, 2, Article 9 (March 2020), 26 pages. https://doi.org/10.1145/3368368

[7] Rocío Calvo, Shaun K. Kane, and Amy Hurst. 2014. Evaluating the Accessibility of Crowdsourcing Tasks on Amazon's Mechanical Turk. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility* (Rochester, New York, USA) *(ASSETS '14)*. Association for Computing Machinery, New York, NY, USA, 257–258. https://doi.org/10.1145/2661334.2661401

[8] Dipanjan Chakraborty, Indrani Medhi, Edward Cutrell, and William Thies. 2013. Man versus Machine: Evaluating IVR versus a Live Operator for Phone Surveys in India. In *Proceedings of the 3rd ACM Symposium on Computing for Development* (Bangalore, India) *(ACM DEV '13)*. Association for Computing Machinery, New York, NY, USA, Article 7, 9 pages. https://doi.org/10.1145/2442882.2442891

[9] Manu Chopra, Indrani Medhi Thies, Joyojeet Pal, Colin Scott, William Thies, and Vivek Seshadri. 2019. Exploring Crowdsourced Work in Low-Resource Settings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300611

[10] Djellel Difallah, Elena Filatova, and Panos Ipeirotis. 2018. Demographics and Dynamics of Mechanical Turk Workers. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (Marina Del Rey, CA, USA) *(WSDM '18)*. Association for Computing Machinery, New York, NY, USA, 135–143. https://doi.org/10.1145/3159652.3159661

[11] Ding. 2020. https://www.ding.com/.

[12] Nathan Eagle. 2009. Txteagle: Mobile Crowdsourcing. In *Proceedings of the 3rd International Conference on Internationalization, Design and Global Development: Held as Part of HCI International 2009* (San Diego, CA) *(IDGD '09)*. Springer-Verlag, Berlin, Heidelberg, 447–456. https://doi.org/10.1007/978-3-642-02767-3_50

[13] Easypaisa. 2020. https://easypaisa.com.pk/.

[14] Hamid Ekbia and Bonnie Nardi. 2016. Social Inequality and HCI: The View from Political Economy. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4997–5002. https://doi.org/10.1145/2858036.2858343

[15] Hamid R Ekbia and Bonnie A Nardi. 2017. *Heteromation, and other stories of computing and capitalism.* MIT Press, Cambridge, MA, USA.

[16] ethnologue.com. 2017. https://www.ethnologue.com/country/pk/languages.

[17] Pamina Firchow and Roger Mac Ginty. 2020. Including hard-to-access populations using mobile phone surveys and participatory indicators. *Sociological Methods & Research* 49, 1 (2020), 133–160.

[18] Freeswitch. 2020. https://freeswitch.com/.

[19] Ujwal Gadiraju, Sebastian Möller, Martin Nöllenburg, Dietmar Saupe, Sebastian Egger-Lampl, Daniel Archambault, and Brian Fisher. 2017. Crowdsourcing versus the laboratory: towards human-centered experiments using the crowd. In *Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments*. Springer, New York, NY, USA, 6–26.

[20] Dustin G Gibson, Adaeze C. Wosu, George William Pariyo, Saifuddin Ahmed, Joseph Ali, Alain B Labrique, Iqbal Ansary Khan, Elizeus Rutebemberwa, Meerjady Sabrina Flora, and Adnan A Hyder. 2019. Effect of airtime incentives on response and cooperation rates in non-communicable disease interactive voice response surveys: randomised controlled trials in Bangladesh and Uganda. *BMJ Global Health* 4, 5 (2019). https://doi.org/10.1136/bmjgh-2019-001604 arXiv:https://gh.bmj.com/content/4/5/e001604.full.pdf

[21] Mary L Gray and Siddharth Suri. 2019. *Ghost work: how to stop Silicon Valley from building a new global underclass.* Eamon Dolan Books, Boston, MA, USA.

[22] GSMA. 2020. https://www.gsma.com/mobileeconomy/wp-content/uploads/2020/03/GSMA_MobileEconomy2020_Global.pdf.

[23] Aakar Gupta, William Thies, Edward Cutrell, and Ravin Balakrishnan. 2012. MClerk: Enabling Mobile Crowdsourcing in Developing Regions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1843–1852. https://doi.org/10.1145/2207676.2208320

[24] Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P. Bigham. 2018. *A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk.* Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174023

[25] Kotaro Hara and Jeffrey P. Bigham. 2017. Introducing People with ASD to Crowd Work. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) *(ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 42–51. https://doi.org/10.1145/3132525.3132544

[26] Lilly C. Irani and M. Six Silberman. 2013. Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 611–620. https://doi.org/10.1145/2470654.2470742

[27] Lilly C. Irani and M. Six Silberman. 2016. Stories We Tell About Labor: Turkopticon and the Trouble with "Design". In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4573–4586. https://doi.org/10.1145/2858036.2858592

[28] Jazzcash. 2020. https://www.jazzcash.com.pk/.

[29] Shashank Khanna, Aishwarya Ratan, James Davis, and William Thies. 2010. Evaluating and Improving the Usability of Mechanical Turk for Low-Income Workers in India. In *Proceedings of the First ACM Symposium on Computing for Development* (London, United Kingdom) *(ACM DEV '10)*. Association for Computing Machinery, New York, NY, USA, Article 12, 10 pages. https://doi.org/10.1145/1926180.1926195

[30] Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing User Studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 453–456. https://doi.org/10.1145/1357054.1357127

[31] Terri Lander, Ronald A Cole, Beatrice T Oshika, and Mike Noel. 1995. The OGI 22 language telephone speech corpus. In *Fourth European Conference on Speech Communication and Technology*. Fourth European Conference on Speech Communication and Technology, Madrid, Spain.

[32] 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* (London, United Kingdom) *(ACM DEV '10)*. Association for Computing Machinery, New York, NY, USA, Article 2, 8 pages. https://doi.org/10.1145/1926180.1926183

[33] UNESCO Literacy. 2020. https://en.unesco.org/themes/literacy.

[34] HLR Lookup. 2020. https://www.hlrlookup.com.

[35] Meghana Marathe, Jacki O'Neill, Paromita Pain, and William Thies. 2015. Revisiting CGNet Swara and Its Impact in Rural India. In *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development* (Singapore, Singapore) *(ICTD '15)*. Association for Computing Machinery, New York, NY, USA, Article 21, 10 pages. https://doi.org/10.1145/2737856.2738026

[36] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods* 44, 1 (2012), 1–23.

[37] Indrani Medhi, Aman Sagar, and Kentaro Toyama. 2006. Text-free user interfaces for illiterate and semi-literate users. In *2006 international conference on information and communication technologies and development*. IEEE, New York, NY, USA, 72–82.

[38] B. Mumtaz, S. Rauf, H. Qadir, J. Khalid, T. Habib, S. Hussain, R. Barkat, and Ehsan-ul-Haq. 2018. URDU Speech Corpora for Banking Sector in Pakistan. In *2018 Oriental COCOSDA - International Conference on Speech Database and Assessments*. IEEE, New YOrk, NY, USA, 9–14.

[39] Bonnie Nardi. 2019. Design in the Age of Climate Change. *She Ji: The Journal of Design, Economics, and Innovation* 5, 1 (2019), 5–14.

[40] Prayag Narula, Philipp Gutheim, David Rolnitzky, Anand Kulkarni, and Bjoern Hartmann. 2011. MobileWorks: A Mobile Crowdsourcing Platform for Workers at the Bottom of the Pyramid. *Human Computation* 11, 11 (2011), 45.

[41] Mustafa Naseem, Bilal Saleem, Sacha St-Onge Ahmad, Jay Chen, and Agha Ali Raza. 2020. An Empirical Comparison of Technologically Mediated Advertising in Under-Connected Populations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376683

[42] Pakistan Bureau of Statistics. 2019. http://www.pbs.gov.pk/content/pakistan-social-and-living-standards-measurement.

[43] Jonas Oppenlaender, Kristy Milland, Aku Visuri, Panos Ipeirotis, and Simo Hosio. 2020. Creativity on Paid Crowdsourcing Platforms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14.

https://doi.org/10.1145/3313831.3376677

[44] George W. Pariyo, Abigail R. Greenleaf, Dustin G. Gibson, Joseph Ali, Hannah Selig, Alain B. Labrique, Gulam Muhammed Al Kibria, Iqbal Ansary Khan, Honorati Masanja, Meerjady Sabrina Flora, Saifuddin Ahmed, and Adnan A. Hyder. 2019. Does mobile phone survey method matter? Reliability of computer-assisted telephone interviews and interactive voice response non-communicable diseases risk factor surveys in low and middle income countries. *PLOS ONE* 14, 4 (04 2019), 1–25. https://doi.org/10.1371/journal.pone.0214450

[45] 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* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 51–54. https://doi.org/10.1145/1518701.1518709

[46] Ellie Pavlick, Matt Post, Ann Irvine, Dmitry Kachaev, and Chris Callison-Burch. 2014. The language demographics of amazon mechanical turk. *Transactions of the Association for Computational Linguistics* 2 (2014), 79–92.

[47] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology* 70 (2017), 153 – 163. https://doi.org/10.1016/j.jesp.2017.01.006

[48] M. Qasim, S. Rauf, S. Hussain, and T. Habib. 2016. Urdu speech corpus for travel domain. In *2016 Oriental COCOSDA - International Conference on Speech Database and Assessments*. IEEE, New YOrk, NY, USA, 237–240.

[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* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174217

[50] 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 *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300642

[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* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2803–2812. https://doi.org/10.1145/2470654.2481389

[52] Recharge. 2020. https://www.recharge.com.

[53] Niloufar Salehi, Lilly C. Irani, Michael S. Bernstein, Ali Alkhatib, Eva Ogbe, Kristy Milland, and Clickhappier. 2015. We Are Dynamo: Overcoming Stalling and Friction in Collective Action for Crowd Workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 1621–1630. https://doi.org/10.1145/2702123.2702508

[54] Samasource. 2020. https://www.samasource.com/.

[55] Saiph Savage, Chun Wei Chiang, Susumu Saito, Carlos Toxtli, and Jeffrey Bigham. 2020. Becoming the Super Turker:Increasing Wages via a Strategy from High Earning Workers. In *Proceedings of The Web Conference 2020* (Taipei, Taiwan) *(WWW '20)*. Association for Computing Machinery, New York, NY, USA, 1241–1252. https://doi.org/10.1145/3366423.3380200

[56] Kerstin EE Schroder, Christopher J Johnson, and John S Wiebe. 2007. Interactive voice response technology applied to sexual behavior self-reports: A comparison of three methods. *AIDS and Behavior* 11, 2 (2007), 313–323.

[57] Selenium. 2020. https://selenium-python.readthedocs.io/.

[58] M. Six Silberman. 2015. Information systems for the age of consequences. *First Monday* 20, 8 (Jul. 2015). https://doi.org/10.5210/fm.v20i8.6128

[59] M. S. Silberman, B. Tomlinson, R. LaPlante, J. Ross, L. Irani, and A. Zaldivar. 2018. Responsible Research with Crowds: Pay Crowdworkers at Least Minimum Wage. *Commun. ACM* 61, 3 (Feb. 2018), 39–41. https://doi.org/10.1145/3180492

[60] Charles Ssemugabo, Elizeus Rutebemberwa, Dan Kajungu, George W Pariyo, Adnan A Hyder, and Dustin G Gibson. 2019. Acceptability and Use of Interactive Voice Response Mobile Phone Surveys for Noncommunicable Disease Behavioral Risk Factor Surveillance in Rural Uganda: Qualitative Study. *JMIR Form Res* 3, 4 (3 Dec 2019), e15000. https://doi.org/10.2196/15000

[61] Saiganesh Swaminathan, Indrani Medhi Thies, Devansh Mehta, Edward Cutrell, Amit Sharma, and William Thies. 2019. Learn2Earn: Using Mobile Airtime Incentives to Bolster Public Awareness Campaigns. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 49 (Nov. 2019), 20 pages. https://doi.org/10.1145/3359151

[62] 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* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 417–426. https://doi.org/10.1145/2702123.2702191

[63] Aditya Vashistha, Abhinav Garg, and Richard Anderson. 2019. ReCall: Crowdsourcing on Basic Phones to Financially Sustain Voice Forums. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300399

[64] Aditya Vashistha, Abhinav Garg, Richard Anderson, and Agha Ali Raza. 2019. Threats, Abuses, Flirting, and Blackmail: Gender Inequity in Social Media Voice Forums. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300302

[65] 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* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1855–1866. https://doi.org/10.1145/3025453.3025640

[66] 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* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173631

[67] Haohan Wang, Agha Ali Raza, Yibin Lin, and Roni Rosenfeld. 2013. Behavior Analysis of Low-Literate Users of a Viral Speech-Based Telephone Service. In *Proceedings of the 4th Annual Symposium on Computing for Development* (Cape Town, South Africa) *(ACM DEV-4 '13)*. Association for Computing Machinery, New York, NY, USA, Article 12, 9 pages. https://doi.org/10.1145/2537052.2537062