Evaluating and Improving the Usability of Mechanical Turk for Low-Income Workers in India

Shashank Khanna
IIT Bombay
shashank.khanna@gmail.com

Aishwarya Ratan
Microsoft Research India
aratan@microsoft.com

James Davis
UC Santa Cruz
davis@cs.ucsc.edu

William Thies
Microsoft Research India
thies@microsoft.com

ABSTRACT

While platforms such as Amazon Mechanical Turk have generated excitement as a potential source of income in developing regions, to date there remains little evidence that such opportunities have transformed livelihoods for low-income workers. In this study, we analyze the usability barriers that prevent those with basic digital literacy skills from accomplishing simple tasks on Mechanical Turk. Based on our observations, we design new user interfaces that reduce the barriers to task comprehension and execution. Via a study of 49 low-income workers in urban India, we demonstrate that new design elements – including simplified user interfaces, simplified task instructions, and language localization – are absolutely necessary to enable low-income workers to participate in and earn money using Mechanical Turk. We synthesize our findings into a set of design recommendations, as well as a realistic analysis of the potential for microtasking sites to deliver supplemental income to lower-income communities.

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]; H.5.4 [**Hypertext/Hypermedia**]; K.4.2 [**Social Issues**]: Employment

General Terms

Design, Experimentation, Human Factors

Keywords

Mechanical Turk, ICT4D, microtasks, India

1. INTRODUCTION

The emergence of microtask marketplaces such as Amazon Mechanical Turk have made it possible for anyone with access to a computer and Internet connection to earn small amounts of money by completing small jobs online. A recent industry report estimates that in the last decade, over 1 million workers have earned $1-2 billion via crowdsourced work allocation [3]. Microtasks posted to Mechanical Turk have an estimated payout of $2,000 per day [4], and 50 other companies are developing online task marketplaces of various kinds [3].

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

Microtasking services may hold particular promise for workers in low-income countries. They allow individuals the possibility of working and earning without the need for physical co-location, pre-existing employment contracts, or even an established identity or history with the employer. The only requirement for being paid is satisfactory completion of the task at hand. This arrangement would seem to lower the barriers to entry and allow a wide range of workers to participate in and benefit from income-generating opportunities. Recent microtasking platforms such as Samasource and txteagle aim to deliver on this promise by specifically targeting workers in developing regions.

However, though microtasking may be perceived as a portal of opportunity for low-income workers, to date there has been little rigorous study of the actual barriers and potential in realizing this opportunity. A prior survey of 733 workers on Mechanical Turk found that 36% were located in India [14]. Indian respondents were highly educated, with 66% having a college degree or higher. While 64% of Indian respondents reported an annual household income of less than $10,000 per year, only 27% reported occasional or regular reliance on MTurk to "make basic ends meet".

In this paper, we evaluate and improve the usability of Mechanical Turk (MTurk) specifically for low-income workers in India. Our target users are those who have finished 10-12 years of schooling and earn less than $1,700 (Rs. 75,000) per year[1]. Such workers find employment in a variety of service sector occupations, including housekeeping and catering staff, drivers, security guards, maintenance crew, retail sales agents, etc. While there are 61 million workers (13% of India's labor force) that match these education and income criteria [7], we found (via a survey of MTurk) that less than 3% of respondent India-based Turkers fall into this demographic.

Our study proceeds in three steps. First, we observe 7 low-income workers attempting tasks on the current version of MTurk. This exposes many usability barriers, spanning the tasks' written instructions, the user interface, and the cultural context. Informed by these observations, we undertake improved designs for MTurk, incorporating simplified instructions and interfaces, language localization, and video tutorials. Finally, we evaluate five alternative designs via a study of 49 workers, in which participants are asked to complete an image-annotation task.

Our basic results are as follows. While there are tasks on MTurk that (if understood) can be successfully completed by low-income workers in India, the current MTurk interface represents a significant barrier in preventing workers from accomplishing such tasks on their own. In the case of image annotation (one of the easiest tasks available), none of our participants were able to annotate a single image using the existing interface. A literal translation into the local language was also insufficient to enable any task

[1]Throughout the paper, we use an exchange rate of 1 US Dollar = 45 Indian Rupees.

completion. However, upon adopting our improved user interface (including simplified task instructions and language localization), participants correctly annotated an average of 66% of the images. Replacing the text instructions with a video tutorial did not significantly affect performance. We formulate our findings as a set of design recommendations, which we hope will enable microtasking to extend its reach to lower-income populations.

To summarize, this paper makes the following contributions:

- A new survey (reaching 200 respondents via MTurk) that deepens our understanding of the demographics and motivations of Turkers in India (Section 3).
- An observation of 7 low-income workers' first-time experience of MTurk in India. Our observations are formulated as a set of usability barriers (Section 4).
- A new design for a microtasking platform, incorporating simplified interfaces and instructions, language localization, and (optionally) a video tutorial (Section 5).
- A user study with 49 low-income workers that measures the impact of various design elements on task completion. Our design recommendations prove vital for enabling satisfactory success rates (Section 6).

Section 7 discusses our overall recommendations, the potential for workers to earn money and develop professionally, and the limitations of our study. We conclude in Section 8.

2. RELATED WORK

The work most closely related to ours are microtasking platforms that specifically target workers in low-income countries. One such service is txteagle, which distributes text-based tasks via SMS for workers to perform on mobile phones in Kenya [2]. Sample tasks include software localization, evaluation of search results, categorization of blog sentiments, and market research; there are also plans to add voice tasks, enabling jobs such as transcription. The target users are similar to those we consider in this paper. An informal study of taxi drivers, security guards, and high school students in Kenya showed that they could complete translation tasks with approximately 75% accuracy, though few details are available [2]. We conjecture that this platform would be more challenging to use than Mechanical Turk (on a computer), because all of the tasks are based on text and also require text entry on a mobile keypad. However, the target users may be more familiar and comfortable with using SMS than they are with using a computer, which could boost their relative productivity on the mobile interface. Moreover, the cost of computer access is eliminated for users who have access to a mobile phone.

Samasource is an nonprofit organization that offers paid microtasks to women, youth, and refugees with the help of local "service partners" that manage the recruitment and training of workers [15]. To clients, they offer services such as business listings verification, audio/video transcription, image tagging, translation, and data entry. They also offer an iPhone application, Give Work, which enables volunteers in rich countries to verify the accuracy of Samasource workers in developing regions (if the volunteer's answer matches the worker's answer, then the worker is paid). While Samasource also targets low-income workers in India, the profiles of 19 India-based workers on their website [15] suggests that the median level of education is a 3-year Bachelor's degree, which is higher than that of our target users. We are unaware of any usability analysis of their platform.

In addition to these services, there are dozens of microtasking platforms that likely draw workers from developing regions [3]. In addition to Mechanical Turk, some examples include CastingWords, CloudCrowd, CrowdFlower, CrowdSifter, DataDiscoverers, LeadVine, LiveWork, LogoTournament, and SmartSheet. Also, a platform called The Extraordinaries represents a micro-volunteering network for the benefit of non-profit organizations.

We are not the first researchers to examine the demographics of users on Mechanical Turk. As detailed in the prior section, Ross et al. conducted an online survey of 733 Turkers [14] and found that 36% of respondents were located in India. Ipeirotis describes a similar survey of 1,000 Turkers [5] and finds 34% to be based in India; respondents from India are compared to those from the United States along various dimensions. While our results are not starkly different from either of these surveys, we focus exclusively on Turkers based in India and customize our questions to the local context (e.g., asking about income in Rupees per month, and education according to the Indian system) which may improve the quality of data. We also probe workers' qualitative motivations and experiences with MTurk.

Outside the context of microtasking, other researchers have also examined the interaction of low-income and low-literate users with computer technology. Medhi et al. develop user interfaces that cater to non-literate and semi-literate users [11], with applications in health [10], job search [8], and mobile banking [9]. While some of the principles espoused – such as using appropriate terminology in the local language – also apply in our scenario, our design task is different in that we target literate users with basic computer skills. Ratan et al. also explore the potential value of free on-site computer access for support staff in urban office facilities in India [13]. While their target user group overlaps with ours, the researchers did not observe or promote usage of microtasking sites during their study period. Such contexts could provide an ideal environment to deploy our designs in the future.

3. STATUS AND POTENTIAL OF MECHANICAL TURK IN INDIA

To better understand the current usage of MTurk in India, we conducted a survey of 200 Indian Turkers in July, 2010. The survey was circulated as a task on MTurk, and requested information on participants' income levels, education, ownership of key assets, their discovery of MTurk, their current usage patterns, and the impact of MTurk on their lives.

We found this user group to be young and even more highly educated than the respondents of Ross et al. [14]. The average Turker respondent from India in this case is 27 years old. Close to 80% of respondents have completed a Bachelor's degree or higher, with another 11% currently in college. 60% have been educated in schools where English is the medium of instruction. 92% of respondents have a PC and Internet connection in their homes. Over 75% own a motorcycle and a refrigerator, 60% own a washing machine, while 28% own a car. The geographic spread of respondent Turkers within India is very wide, including those from metros like Chennai, Kolkata and Bangalore, as well as those who live in smaller cities and towns like Erode, Allepey, Guntur, Hisar and Bhatinda.

The median Indian Turker respondent reported an individual income of $2700 annually and an annual household income of twice that amount ($5300). Median earnings from MTurk are $8 per week. What is significant though is that this amount constitutes 15% of the median Turker's total individual income. Nearly a quarter of the respondents earn >$65 per month from Turk. It is clear that for those with lower income levels, the MTurk component of

their earnings plays an important role in their economic welfare. Unsurprisingly, 88% of our Indian Turker respondents indicate that earning income is their primary motivation for using MTurk, in contrast to 8% for whom this is primarily an activity for entertainment.

The importance of MTurk in a number of respondents' lives is illustrated by the responses to the open-ended questions we asked in the survey. A 26-year old college graduate from Kolkata relies on it as his primary income source, earning $155 from MTurk work per month: "I'm from a middle class family. After completing my degree I looked for job everywhere but failed. But when I found mturk, it changed my life. It helped me a lot." Several respondents ask for more work to be available to them on this platform. A 31-year old college graduate from Coimbatore currently earns all of her $67 of income from MTurk each month, which is a quarter of her entire household's income. She views this as an important component of saving for her children's education: "please give me a job (easy and more pay) continuously so that to earn little bit more amount to manage my house expenses and to save for the studies of the 2 children." Income itself is an intermediary outcome toward fulfilling real capabilities such as learning and self-reliance, as a respondent from Trichy earning $133 a month on Turk writes: "MTurk really an advantage to me, it helps me to pay my college fees myself. It made me to feel I'm on my own. I got the respect while studying by this reasonable income."

Untapped potential: Our survey indicates that while MTurk plays a substantial role as an income-generator among our Indian Turker respondents, relative to other wage earners in India, those using MTurk come from relatively well-off settings. Those with a Bachelor's degree or higher constitute only 6% of India's working age population (15-60 years) [7]. Similarly, home PC penetration in India is estimated at <10% [6]. For MTurk to play a more substantial role as a supplemental income generator, it would need to allow larger numbers of individuals from marginally lower-income and lower-education backgrounds to participate. With 75% of secondary school graduates (class 12) in India earning less than $1100 a year [7], there is immense scope for even small supplemental earning streams to have a significant impact on individual and household economic well-being for a wider share of the population. There has been no study thus far to assess whether the barrier to participation in microtasking marketplaces for this segment of workers is access, ability, language, or some combination of these and other factors. We use this open question as a starting point to understand whether there is scope for lower-income workers in India to benefit from income-earning opportunities on the Web.

4. USER OBSERVATION

In the first phase of our investigation, we aimed to observe potential Turkers from our target segment and understand their experience as they interacted with the current version of MTurk. The barriers observed during this phase served to inform our own design, which is detailed in the next section.

4.1 Participants

Our focus group consisted of seven office support staff from the low-income segment described previously, spanning security guards, housekeeping staff, maintenance staff, a driver, and an IT assistant. They had received an average of 11 years of education. While they all had some familiarity with English, generally their schooling was in a local language. They had acquired basic IT skills via on-site access to a computer, as well as other sources. None had previously used MTurk.

Task	Input Method	Output Method	Instances Tested
Image labeling	Graphical	Graphical	4
CAPTCHA decoding	Graphical	Text	4
Address verification	Text	Text	3

Table 1: MTurk tasks used for observation sessions.

4.2 Methodology

To understand the barriers and potential of the existing MTurk platform as an income-generating opportunity for our target users, we conducted informal sessions in which we observed (and also coached) users during their first interaction with MTurk.

We conducted an hour-long, one-on-one session with each participant. We started by giving a basic explanation of MTurk and the context of the study. Then, we asked participants to register a new account on the MTurk website, and to complete 1 or 2 pre-selected tasks (time willing). While we did not offer any demonstration of the tasks up-front, we did prompt users with assistance if they became stuck. We utilized the unmodified MTurk interface (in English) throughout the study.

We selected three tasks for the study (see Table 1). As the tasks on MTurk can be broadly categorized as either text or image-based, we aimed for a selection of tasks that had varying dependence on either images or text. In the image labeling task, users were asked to label images in various ways, including drawing a bounding box around each human, verifying the bounding boxes of others, or selecting those images containing a given object (a fish). In the CAPTCHA task, users had to type the distorted letters appearing in experimental CAPTCHAs [17]. In the address verification task, users were asked to utilize Internet search to verify that a bank had a given mailing address.

4.3 Results

Overall, participants had considerable difficulty using the MTurk interface in its current form. Despite extensive prompting, no participants were able to complete the text-based task (address verification). While all participants eventually completed the image labeling tasks, and 75% of participants completed the CAPTCHA tasks, these also required extensive prompts, and often resulted in incorrect answers. As the purpose of this initial investigation was to observe the barriers to using MTurk, we focus on the qualitative experience gained and postpone a rigorous analysis of accuracy and task completion to Section 6. We separate our discussion of the usability barriers into two categories: general and task-specific.

General Barriers

As a whole, the MTurk site posed several difficulties for the participants. These included:

- **Complexity of instructions.** As non-native speakers of English, participants were often unable to understand the detailed (and somewhat ad-hoc) instructions that accompanied each task (e.g., see Figure 1). For example, the word "keyword" was unfamiliar, and interpreted as something that was typed with the keyboard. In lieu of understanding all of the written text, several participants focused their attention on any images (or highlighted words) in the instructions, though typically these did not convey all of the information needed. On occasions when participants seemed to give up on fully understanding the instructions, they would click on nearby links (for example, titled "Go"), though these often led to unexpected behavior.

- **User interface complexity.** Participants were often overwhelmed and confused with the number of buttons on the screen. There was a lack of separation between task-specific functionality and general MTurk functionality (presented on the MTurk dashboard at the top), making it difficult to restrict attention to the task at hand. This problem is illustrated in Figure 3.
- **Navigation difficulties.** Several users lost all of their work on a task because they accidentally pressed the backspace key (without focusing on a text box), causing the browser to go back to the prior page, which was the instructions or registration page. Some users also pressed the back button explicitly, causing the same effect. The MTurk window also contained nested scrollbars: one for the MTurk page, and one for the enclosed task page. Participants often used the wrong scrollbar, missing important task content that was accessible via horizontal or vertical scrolling in the inner window.
- **Sequencing problems.** More than one user started working on a task without realizing he needed to "accept" it, making that work ineligible for payment. At the same time, detailed instructions for a task came only after the user officially accepted it, and instructions did not remain available for future reference.
- **Cultural context.** Occasionally cultural differences made it more difficult for participants to complete the task. For example, one participant did not recognize a Western-style kitchen, because it lacked vegetables and a gas stove; he guessed that it was a bedroom or drawing room instead. On the registration page, some options are not tuned for the Indian context; for example, "Class 10" is not listed as an option for education, and Kannada is missing from the list of languages.

Task-Specific Barriers

In addition to the general barriers to using the MTurk site, we also outline the more specific difficulties that inhibited participants from completing each task.

The address verification task required nuanced Internet search skills that were beyond the training and experience level of our participants. To start, participants had difficulty understanding the intent of the task. However, even upon gaining an understanding based on our prompts, it was difficult for participants to distinguish trustworthy from non-trustworthy sources of information on the Internet. This is a difficult task for even experienced Internet users. Participants often took their answers from the initial page of results from a search engine (without clicking on any target pages), which was generally insufficient to find the information needed. We believe that this task may be beyond reach of novice computer and Internet users, regardless of the interfaces developed.

For the CAPTCHA tasks, one participant read letters from top-to-bottom, rather than left-to-right. There was some confusion between lower- and upper-case letters. One participant entered spaces between each pair of letters; another asked how to enter slanted characters. There was basic difficulty navigating the cursor to correct mistakes in text boxes. When there was no visible text in an image, one user became stuck and took too much time before moving on. While some of these mistakes could perhaps be avoided by giving more example CAPTCHAs in the instructions, this task still requires users to deal with the ambiguity of whether it is humanly possible to decode the printed letters. We speculate that this may represent a barrier to efficient and confident task completion regardless of the interface design.

		Original Interface		New Interface
		English	Kannada	Kannada
Original Instructions	English text	X		
	Kannada text		X	
New Instructions	Kannada text			X
	Kannada video	X		X

Table 2: Summary of the designs implemented.

For the image labeling tasks, participants were largely successful after gaining a baseline understanding of the user interface and the expectations of the task. Some participants were unfamiliar with the click-and-drag interaction style needed to draw a bounding box; one participant used the wrong mouse button at first. Some prompts were needed to clarify the terms used to assess others' results ("good", "bad", or "good with errors"). One user entered an assessment of the overall quality of an image, rather than the quality of the bounding boxes drawn. One user drew a bounding box around two humans, rather than drawing two separate boxes, as this was not explicitly prohibited by the instructions.

Based on the experiences above, we deemed the image labeling tasks to be the most accessible to low-income workers. While extensive prompts were needed to coach participants through the tasks, once an understanding was gained, they were generally able to complete the job. Our goal in the remainder of the paper is to see if this understanding can be imparted via a standalone, unsupervised system, thereby making the system accessible to the broader populace.

5. DESIGNING A SOLUTION

In order to address the usability barriers observed in the prior section, we explored three dimensions of the design space:

1. **Improved interface and instructions.** We redesigned the user interface and re-wrote each task's instructions to be more accessible to novice users. Our design guidelines are detailed in this section.
2. **Language localization.** Rather than using English, we translated[2] the instructions and interfaces to the local language (Kannada).
3. **Video tutorials.** We also experimented with a video tutorial, whereby users are introduced to the task by a two-minute video demonstration (recorded as a screencast). This idea was inspired by the prior success of "full-context videos" for first-time computer users [12]. The narration of the video is similar to the written instructions, but the text instructions are not shown.

These axes define a rich design space, out of which we implemented a selection of five designs (see Table 2). Two of the designs represent baselines, utilizing the original interface and instructions in either English or Kannada. Two of the designs represent new best-effort systems, utilizing a new interface and new instructions (in either text or video format). The last design represents an intermediate point, utilizing the original English interface but with a

[2]The Kannada version of the original interface was obtained via professional, word-for-word translation of the English text. However, the Kannada version of the new interface was composed from scratch in Kannada, leading to more simple language throughout.

Figure 1: Original instructions for bounding box task.

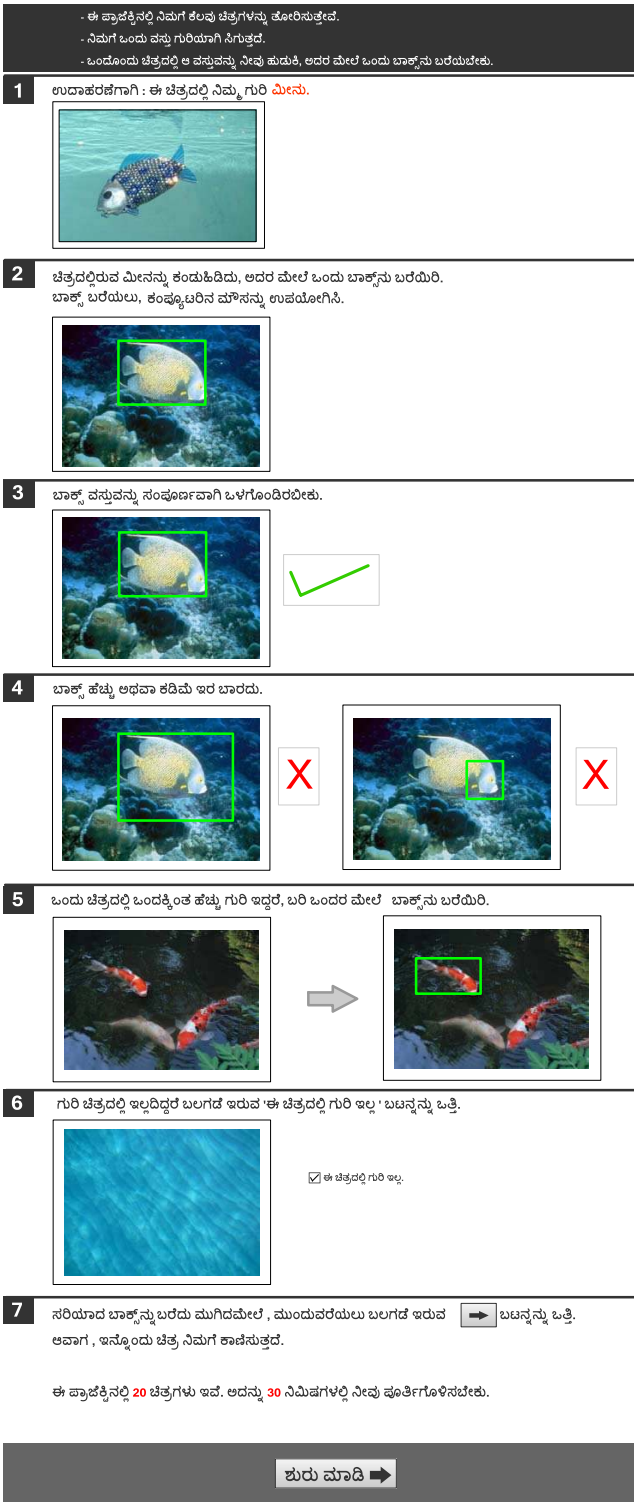

Figure 2: New instructions for bounding box task.

Kannada video tutorial, to see if this is sufficient to enable workers to complete the tasks.

The most innovative and generalizable elements of our design are embedded in the new user interface, with accompanying instructions and illustrations for each task. An example of this design appears in Figures 2 and 4, while the original design is available for comparison in Figures 1 and 3. Our design adheres to the following guidelines, each of which corresponds to a usability barrier experienced by participants in our observation sessions:

- **Use simple, illustrated instructions for each task.** The instructions for tasks on MTurk are often written in an unstructured way and rely on complex, ad-hoc figures (see Figure 1). We invested significant effort into distilling the instructions into a series of clear, illustrated steps. As shown in Figure 2, each step is numbered and includes a graphical example. Be-

Figure 3: Original interface for bounding box task.

cause the instructions are written in Kannada, we provide a partial translation below:

> In this project we will show you some pictures.
> You will get a target object.
> In each picture, you should search for that object and draw a box around it.
>
> 1. For example: In this picture, your target is fish.
>
> 2. Search and find the fish in the picture, and then draw a box around it. To draw the box, use the computer's mouse.

This format is similar to what one might expect to find in airline safety pamphlets, fire extinguishers, and other critical contexts where the instructions cannot afford to be misunderstood by anyone in the world. In some cases, we included both positive and negative examples of how to handle various cases, as well as how to deal with unexpected or unusual instances of tasks.

- **Minimize visual complexity.** Unlike the original MTurk interface, which includes a complex banner of links unrelated to the task (see Figure 3), our interface focuses the user's attention by eliminating all superfluous functionality (see Figure 4). If the user wants to abort the task, they can navigate with the back button to where they started.

Figure 4: New interface for bounding box task.

- **Streamline navigation.** When the user is in the process of completing a task, they should have only two options: to proceed to the next task instance, or to return to the prior instance. The nested scrollbars of MTurk are eliminated to ensure that the user explores the entire task using only the browser's scrollbar. The back button (or the backspace key, if pressed outside a text box) returns only to the prior instance of a task, rather than resetting the entire process.

- **Anticipate sequencing of steps.** While MTurk requires a user to accept a task prior to reading the detailed instructions, our design follows a more intuitive sequencing. First the user sees an overview of the task, which is followed by the detailed instructions.

6. USER STUDY

The goal of our user study is to evaluate each of the designs described in the prior section. In particular, we are interested in whether a shift from English to Kannada language, from the original to the new interface, and from text-based to video instructions can impact the ability of low-income workers to successfully complete tasks in an unsupervised environment.

6.1 Participants

We conducted our user study with 49 participants across two locations in Bangalore (see Table 3). The first location was an office facility in which the support staff had learned basic digital literacy skills through a previous PC access program. The second was a non-profit IT training centre that offered subsidized courses in Windows, Office, the Internet, desktop publishing and accounting software for individuals from low-income backgrounds.

The average age among participants was 23 years. 55% of participants were men. The group included those who had full-time employment and those who were students or were unemployed. Of participants who were employed, the median income was ~$1330 per year. This places him or her roughly in the second quintile (20%-40%) of the urban earnings distribution[3]. These workers are therefore not drawn from the poorest segments of the Indian workforce, but earn wages that fall in the bottom half of the urban wage distribution, well below the earnings level of current Indian MTurk users.

The average educational attainment among participants was 12 years, corresponding with completion of higher secondary school or Class 12 (min 7 years, max 15 years). 77% had been educated in a local language school. All participants had formally or informally learned basic digital literacy skills, i.e., they were able to use a PC, navigate through the Internet, use Office applications, etc. All except three had no prior exposure to MTurk. The remaining three had participated in our observation sessions, but had attempted CAPTCHA or address verification tasks rather than image annotation.

6.2 Methodology

To evaluate our designs, we utilize an image annotation task that is among the most common of all tasks on MTurk. The task requires users to draw a bounding box around a particular object in an image. If the object does not appear in the image, then this is indicated via a checkbox. In our study, the object of interest is a lamp (see Figure 4) while the instructions use an example of a fish (see Figure 2). The results of such tasks are useful for building training sets for computer vision algorithms; workers are typically paid $0.05 for completing 20 images. There are usually hundreds (sometimes thousands) of instances of bounding box tasks on MTurk at any given time.

Our experiment represents a between-subjects design, in which each group of participants completes a set of tasks using one of the five designs from Table 2. Each group had 10 participants, with the exception of one group, which had only 9 participants (this group evaluated the original MTurk interface, in English).

The study was done in both of the locations from which participants were drawn: an urban office facility and an IT training centre in Bangalore. The study was administered in small groups of 1-5 people at a time. As in our observation sessions, we started by providing each participant with a brief overview of MTurk and the purpose of the study. However, in this case, we did not provide any prompts or assistance to the participants once they had started reading the instructions for the task. (We are interested in the first-time experience of untrained users so as to maximize the potential scale of microtasking in India. While organized training could no doubt overcome some of the issues we observed, this would start to overlap the functions of a BPO.) We also timed each participant, giving him or her a maximum of 30 minutes to complete the task. We recorded screen captures and (for some participants) video recordings during the study.

[3]Calculated in terms of 2004-2005 constant prices using the reference distribution from [7, p.35] and inflation data from [16].

	Average annual income ($)	Average length of education (years)	N*
Support staff at office facility	2,093	11	10
Employed learners at IT training centre	1,240	13	14
Students/unemployed learners at IT training centre	0	12	18

** There were 49 participants. The table omits 3 office staff and 4 trainees, who did not report education or income.*

Table 3: Participants in our user study.

Correctness of the participants' work was subsequently judged via inspection of the screen captures. We consider a bounding box to be "correct" if it encloses the object of interest while being neither too big nor too small. A box is deemed too big if there is more than 10 pixels of space between any edge and the enclosed object. A box is deemed too small if the object protrudes more than 3 pixels beyond any edge of the box. While we do not have precise data on which boxes are accepted or rejected by MTurk requesters, these guidelines correspond to our own intuitions regarding which boxes are accurate and useful.

6.3 Results

Results of the user study are summarized in Table 4 and illustrated graphically in Figure 5. A participant's "score" is the fraction of the 20 assigned images that they annotated correctly. Our results can be summarized as follows:

- Using the original MTurk interface and instructions, none of the participants are able to draw even a single bounding box correctly. This holds true even after translating the interface and instructions into the local language.
- Introducing a video tutorial (narrated in Kannada, but using the original English interface) raises participants' average score to 40% (min 0%, max 70%), which is significantly higher than the original average of zero ($p < 0.001$).
- Utilizing our new design for the interface and instructions significantly improves task completion. Participants' average score using the new interface is 66% (min 20%, max 90%), which is significantly higher than the original interface ($p < 0.001$) and the original interface with video tutorial ($p = 0.03$).
- Replacing the text instructions with video instructions, while using the new interface, does not have a significant effect on task completion. With the video tutorial, participants' average score using the new interface is 63% (min 35%, max 90%).

	Original Interface			New Interface	
	Original Text Instructions		New Video Instructions	New Text Instructions	New Video Instructions
	English	Kannada	Hybrid*	Kannada	Kannada
Average score	0%	0%	40%	66%	63%

** English interface, Kannada video instructions*

Table 4: Results of our user study, in which each participant was assigned 20 instances of a bounding box task. A participant's score represents the fraction of task instances that he or she performed correctly.

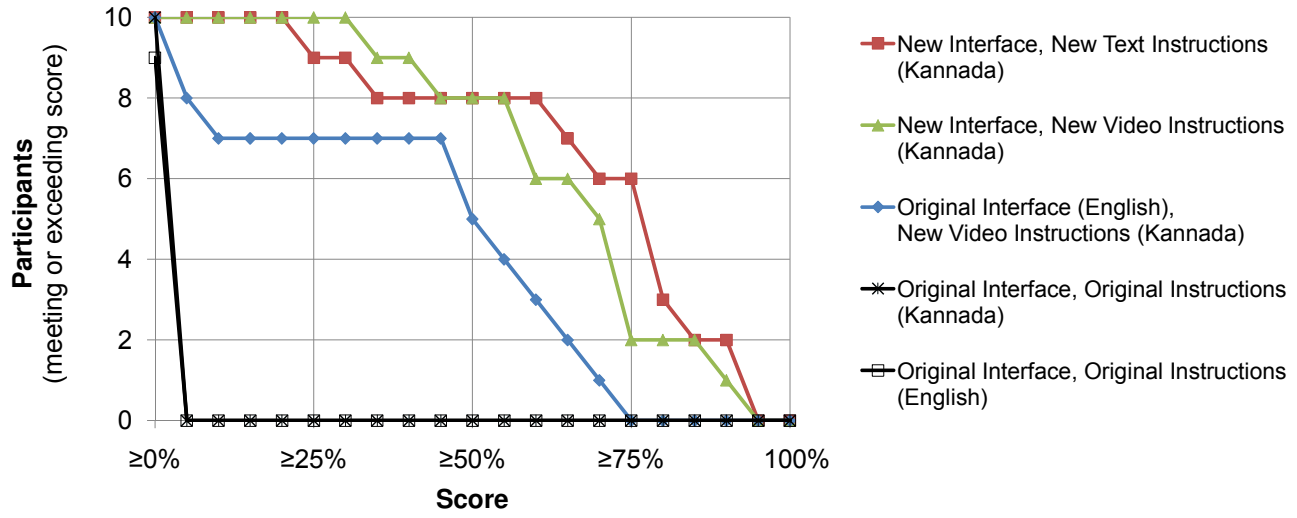

Figure 5: Distribution of scores according to user interface and instructions. The graph shows the number of participants who obtained a given score *or higher*. For example, the number of participants who scored 75% or higher is eight: six on the new interface with text instructions, and two on the new interface with video instructions. Lines closer to the top-right corner are better.

What prevents workers from achieving higher scores? When using our new interface, participants attempted every instance of a task, which implies that low scores are due to incorrect responses. The largest category of errors (amounting to 19% of total errors committed using our new interface with text or video instructions) are due to either annotating a lamp (the object of interest) in a place where no lamp was present, or in asserting that no lamps appeared in the given image when in fact they did. We attribute these errors in part to differences in cultural context, as there were some images of luxurious hotels or Western-style bedrooms in which lamps may not have been recognizable by the participants. In addition, these errors reflect the intrinsic difficulty of the task. Many lamps were partially occluded, omitted, or poorly lit; in fact, one of the authors mistakenly outlined a vacuum cleaner rather than a lamp, and did not notice a lamp attached to a ceiling fan, when performing the task himself.

The next most common category of errors are bounding boxes that were drawn too large, accounting for 11% of the total incorrect instances. Such errors could perhaps be prevented by introducing a "training stage" in which workers' answers are checked on example instances; such training has recently been added for some of these tasks on MTurk (but was not available at the time of our study).

The final category of errors (4.5% of instances) are images that were skipped entirely: participants did not draw a bounding box and also did not check a box to indicate an absence of the target object in the image. We attribute this error to an accidental double click on the button that advances to the next task. It would be straightforward to prevent this behavior by improving the user interface: the user should only be allowed to advance when they have either drawn a box or asserted that no box should be drawn.

7. DISCUSSION

7.1 Design Recommendations

The most striking result of our study is that there exist tasks on MTurk for which the primary barrier to low-income workers is not the cognitive load of the work itself; rather, workers are unable to understand and navigate the tasks due to shortcomings in the user interface, the task instructions, and the language utilized. When we remove these barriers via introduction of improved interfaces and instructions – both in the local language – we observe a dramatic increase in the average participant score from zero to 66%.

Our design recommendations are straightforward. In order to enable successful task completion by low-income workers:

- **Improved interfaces and instructions are needed.** We recommend following the detailed guidelines in Section 5 in order to design a suitable platform for low-income workers. In addition, we suggest including a training and validation stage as part of the task instructions, and performing basic checks (such as non-emptiness) as workers submit each part of a task.
- **Translation to the local language is necessary but not sufficient.** Even following translation, we observed zero task completion with the original interfaces and instructions.
- **Video tutorials are unlikely to be worth the effort.** Replacing the text instructions with video tutorials did not significantly impact task completion (the average score was lower). Given that the videos require considerable effort to produce, this suggests that they are unlikely to be worth it.

Participant	Time required to submit 20 images	Payment for 20 images	Expected hourly payment	Net hourly earnings (with low cost of access)*	Net hourly earnings (with high cost of access)**
Fastest participant (with score ≥ 75%)	1:32	$0.05	$1.96	$1.74	$1.52
Median participant (with score ≥ 75%)	7:20	$0.05	$0.41	$0.19	−$0.03
Slowest participant (with score ≥ 75%)	23:49	$0.05	$0.13	−$0.09	−$0.31

** $0.22 (Rs.10) per hour of PC+Internet access*

*** $0.44 (Rs. 20) per hour of PC+Internet access*

Table 5: Participants' estimated earning potential in completing bounding box tasks on Mechanical Turk.

7.2 Workers' Earning Potential

As many variations of the bounding box task are posted on MTurk at regular intervals, it is a reasonable candidate for sustained opportunity and earnings. We explore this potential by comparing the payment rate expected from MTurk (at the productivity and efficiency levels observed during our user study) to the baseline wages of workers.

One challenge in assessing earning potential is to determine which workers would actually be paid if they had submitted the tasks on MTurk. As requesters on MTurk do not publish the minimum score needed for payment, for the sake of the following analysis we use a threshold score of 75% (that is, any worker who successfully completes at least 75% of the bounding box instances receives full compensation for that task.) As illustrated in Figure 5, there are 8 participants who meet this threshold.

As illustrated in Table 5, estimated earnings from the bounding box task for those who successfully completed it range from $0.13 to $1.96 per hour, depending on the speed of the worker. The fastest worker would earn more than twice the average baseline wage rate ($0.83 per hour) for the employed participants in our study (n=24). However, the median worker would earn $0.41 per hour, which is less than half the baseline wage rate. If the median worker was to spend an hour each day performing such tasks on MTurk at observed efficiency levels, he or she would likely earn $12 per month. This amount represents a 9% income increment compared to the average employed participant's baseline earnings.

To understand the net benefit to the worker, it is also important to consider costs that might be incurred if computer and Internet access are not freely available. If the median worker had to access MTurk at a public shared access point like a cyber café, at a low cost of $0.22 (Rs. 10) per hour, his or her potential earnings from performing similar tasks would halve to $6 per month. It is obvious that the higher the likelihood of workers accessing a PC+Internet facility for free, either at the workplace or at a library or at a friend or relative's home, the higher the share of their MTurk earnings they get to retain.

While the earnings derived for the median worker in our study remain relatively modest, we emphasize that this is not an upper bound on his or her ultimate earning potential. For repeated tasks such as the bounding box, workers may improve their speed over time. Also, there is further room for improvement in our user interface, using mechanisms such as a training phase and input validation as described previously. We consider it to be an interesting research challenge to further improve the hourly wage that is achievable by low-income workers on MTurk.

7.3 Workers' Professional Development

Even if workers can earn money on MTurk, the question remains whether long-term usage of MTurk is beneficial for workers' careers. Zittrain argues that microtasking sites can present many hazards to workers, including the absence of an affiliation with a team, the inability to understand the ultimate applications of one's work (and whether they are morally acceptable), and an absence of regulatory measures such as minimum wage and maximum working hours [18]. A related concern is that microtasking sites might stifle the agency, creativity, or professional development of workers; can workers develop higher-level skills if their work is confined to small, tedious tasks that have no connection to a broader project?

These are important questions that should be considered carefully. Our perspective in this paper has been to explore the potential of microtasking for *supplemental* income generation, which perhaps ameliorates some concerns relative to a full-time undertaking. In fact, one of the benefits of microtasking sites is the flexibility offered: workers can freely experiment with the platform, at any time and for any duration, in a way that is rarely possible with formal employment. For low-income workers who often keep unusual hours, or have unpredictable schedules, this flexibility is the main benefit that we see in exchange for the unregulated working environment.

With respect to skills development, we do believe that microtasking sites could have a lot to offer low-income workers. Computer experience by itself is an important stepping stone for many workers; one could think of MTurk as "computer practice" that is also goal-oriented and income-generating. Moreover, the tasks accessible online are not limited to simple assignments such as drawing bounding boxes; researchers have also utilized MTurk for high-level tasks such as copyediting [1]. Other microtasking sites (such as CrowdSpring and DesignCrowd) leverage workers' creativity to design logos for companies, write taglines for products, and create graphics for brochures or websites. As workers gain experience, they could potentially move to professional sites such as vWorker or Elance, which offer large-scale programming and design projects. Our results indicate that the ease with which workers discover new tasks and opportunities, and skill themselves appropriately through practice sessions or tutorials, is itself contingent on the way the microtasking platform is designed.

7.4 Limitations and Future Work

This study presents learnings from an early exercise on understanding the scope for MTurk to include a wider section of the workforce in countries like India where the need for supplemental earning opportunities is immense.

Our survey of Turkers in India was distributed as a task on MTurk, and is thus prone to self-selection bias. Ross et al. found that their survey respondents were more likely to focus on survey tasks than other tasks [14], which differentiated them from other Turkers. It is possible that the Turkers in our survey also displayed characteristics correlated with an inclination to answering surveys that are more or less generalizable to the entire Indian Turker population.

While the strength of this study lies in its focus on a specific task as an example, clearly we must examine a range of tasks that

meet the skills of our target user group and evaluate performance and earning capabilities across them to produce reliable estimates of average impact. This will include exploring how the UI design and language localization principles that we recommend are optimally implemented at low cost across a variety of tasks and task specification scenarios.

Our study was implemented as a usability evaluation in this instance. Delving into the "practice" of MTurk among our target users in an everyday setting will help us evaluate the other factors that augment or confound usage of this service and its translation into improved earnings. These will include the availability of time to invest in MTurk work and how this varies by age, occupation, and gender; variations in infrastructure availability and reliability that affect users' ability to perform MTurk work effectively; and social interactions that shape MTurk discovery and usage. These questions are best explored as ethnographic studies involving long-term repeated interaction with a select set of individuals from our target user group. Efforts are also underway to deploy our improved designs in an unsupervised real-world setting with monetary compensation for completed tasks, so that we can test whether usage is sustained over time.

8. CONCLUSIONS

While microtasking sites such as Mechanical Turk have succeeded in attracting workers from India, to date they have enrolled very few workers from the low-income segment of the population. In this paper, we offer one of the first analyses of the barriers preventing low-income workers in India from participating in this marketplace, as well as solutions to overcome those barriers.

The first barrier is that of the tasks themselves: the tasks on MTurk require widely varying skills, and certain tasks are beyond the reach of those with limited education and computer experience. For example, we found the task of address verification (via Internet search) to be too difficult for the participants in our observation sessions; we conjecture that many tasks based purely on reading and writing English, or on Web browsing skills, may be a mismatch for the capabilities of our target group. However, importantly, we found graphical tasks such as image annotation to be tractable for low-income workers. The varying skill sets demanded by microtasks suggest that better mechanisms are needed to match tasks with the capabilities of workers.

The second and more interesting barrier that we discovered is unrelated to the tasks themselves, but relates instead to the design of the MTurk website. Even if workers are capable of doing a task, in the existing MTurk platform they are unable to complete the process due to complexities in the user interface as well as the task-specific instructions. These obstacles prevent participants from completing a single image annotation task on the existing MTurk interface – even after it is translated to the local language. This usability challenge should be an eye-opener for microtasking platforms that aim to reach low-income workers.

In order to address some of these usability barriers, we designed a new system that simplifies the user interface as well as the exposition of task-specific instructions. Utilizing our design (which was also translated into the local language), we demonstrate that workers can, on average, successfully perform 66% of the target tasks. Replacing text instructions with video instructions does not significantly affect the rate of task completion.

Our main recommendation for microtasking marketplaces is to clarify the user interface and task-specific instructions according to the guidelines that we have described in Section 7.1. This has the potential to make such platforms accessible and usable to a much broader base of low-income users.

9. ACKNOWLEDGMENTS

We are very grateful to our collaborators at the Hope Foundation, in particular Joseph Kumari, Augustine Thomas, Mini and Sunita, as well as all those from the Shivajinagar center who volunteered to participate in our study. We thank the participants from the support staff of Microsoft Research India, who generously donated their time for the study. We are also grateful to H.G. Srinivasa Prasad and Rashmi K.Y. for help with translation, and to Ed Cutrell and the anonymous reviewers for helpful feedback.

10. REFERENCES

[1] M. Bernstein, G. Little, R. Miller, B. Hartmann, M. Ackerman, D. Karger, D. Crowell, and K. Panovich. Soylent: A word processor with a crowd inside. In *Symposium on User Interface Software and Technology*, 2010.

[2] N. Eagle. txteagle: Mobile Crowdsourcing. In *Internationalization, Design and Global Development*, 2009.

[3] B. Frei. Paid Crowdsourcing: Current State & Progress towards Mainstream Business Use. Smartsheet White Paper, Sep 2009.

[4] P. Ipeirotis. Mechanical turk: Profitable or not? `http://behind-the-enemy-lines.blogspot.com/2009/03/mechanical-turk-profitable-or-not.html`, Mar 2009.

[5] P. Ipeirotis. Demographics of Mechanical Turk. NYU Working Paper No. CEDER-10-01, Mar 2010.

[6] Screen Digest. Market Intelligence Report. 2008.

[7] TeamLease Services. Indian Labour Report 2007: The Youth Unemployability Crisis. `http://www.teamlease.com/images/reports/Teamlease_LabourReport_2007.pdf`, 2007.

[8] I. Medhi, G. Menon, and K. Toyama. Challenges of computerized job-search in the developing world. In *CHI*, 2008.

[9] I. Medhi, G. Nagasena, and K. Toyama. A Comparison of Mobile Money-Transfer UIs for Non-Literate and Semi-Literate Users. In *CHI*, 2009.

[10] I. Medhi, A. Prasad, and K. Toyama. Optimal audio-visual representations for illiterate users. In *World Wide Web Conference*, 2007.

[11] I. Medhi, A. Sagar, and K. Toyama. Text-free user interfaces for illiterate and semiliterate users. *Information Technologies and International Development*, 4(1), 2007.

[12] I. Medhi and K. Toyama. Full-Context Videos for First-Time, Non-Literate PC Users. In *Information and Communication Technologies and Development*, 2007.

[13] A. L. Ratan, S. Satpathy, L. Zia, K. Toyama, S. Blagsvedt, U. S. Pawar, and T. Subramaniam. Kelsa+: Digital Literacy for Low-Income Office Workers. In *Information and Communication Technologies and Development*, 2009.

[14] J. Ross, L. Irani, M. S. Silberman, A. Zaldivar, and B. Tomlinson. Who are the Crowdworkers? Shifting Demographics in Amazon Mechanical Turk. In *alt.CHI*, 2010.

[15] Samasource website. http://www.samasource.org/.

[16] The World Bank. World Development Indicators. `http://data.worldbank.org/indicator/FP.CPI.TOTL.ZG`, 2010.

[17] L. von Ahn, M. Blum, N. J. Hopper, and J. Langford. CAPTCHA: Using Hard AI Problems for Security. In *EUROCRYPT*, 2003.

[18] J. Zittrain. Ubiquitous human computing. *Philosophical Transactions of the Royal Society*, 366, 2008.