Exiting the Design Studio: Leveraging Online Participants for Early-Stage Design Feedback

Xiaojuan Ma[1], Yu Li[2], Jodi Forlizzi[3], Steven Dow[3]
[1]Huawei Noah's Ark Lab, Hong Kong; [2]Cornell University, USA;
[3]Human-Computer Interaction Institute, Carnegie Mellon University, USA;
[1]xiaojuan.c.ma@gmail.com; [2]ly257@cornell.edu; [3]{forlizzi, spdow}@cs.cmu.edu

ABSTRACT
Online collaboration tools enable developers of interactive systems to quickly reach potential users for usability testing. Can these technologies serve designers who seek feedback on user needs during the earliest stages of design? *Online needfinding* may help designers create products and services that can target a more diverse user population. To explore this, we conducted a feasibility study to compare face-to-face methods with online needfinding sessions. We found that video can sufficiently capture nuanced reactions to preliminary concept storyboards, but that feedback providers need guidance and structure. We then introduce a tool for collecting early-stage design feedback from online participants and conduct a case study with a professional design team. The team conducted needfinding activities with local participants, as well as a cost-equivalent number of online participants The case study demonstrates that combining online crowdsourcing with a video survey tool provides a simple and cost-efficient way to collect early-stage feedback.

Author Keywords
Design feedback; needfinding; collaboration; crowdsourcing; storyboard scenarios; online video communication.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Designers interact with potential users through a variety of activities, from concept exploration to usability testing, throughout the design process [22]. Tools for online communication and crowdsourcing have made it possible to conduct some of these interactions online. By making participant recruitment easier, faster, and more diverse, the CSCW research community has started to explore how to leverage online participants to shape the design of interactive systems [14, 18, 3, 19, 32]. For example, Martin et al. show that remote asynchronous usability evaluation reveals similar numbers of problems, faster and cheaper than in-person usability testing [23]. This is important as it helps designers discover usability issues before over-investing in development efforts. Further, designers may reach a more diverse sample of potential users online, which is particularly relevant for practitioners designing software intended for millions of people around the world [9].

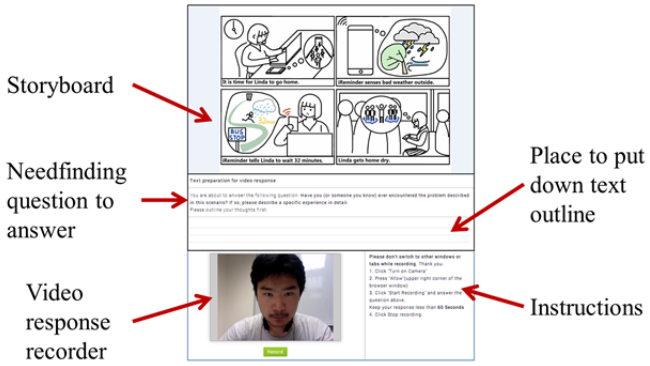

Figure 1. Interface for collecting early-stage design feedback

While online approaches provide a nice option for formative software testing, it remains an open research question whether this translates to the earliest stages of design where the goal is to identify the right problem to solve [16, 28, 33]. Traditional needfinding activities such as "speed dating" with concept storyboards [12] help designers discover the tacit needs and opportunities that ultimately shape a design. At first blush, online strategies seem amenable to helping designers uncover a more diverse set of considerations. However, with in-person needfinding, designers must carefully interpret users' subtle, often non-verbal, reactions to novel interaction scenarios [12]. This issue may make it difficult to conduct needfinding online.

Research by Bietz et al. shows the importance of communication medium when exchanging feedback online; without audio-visual information, creators interpret feedback to be more negative and less credible, and they are less likely to incorporate feedback [7]. The choice of medium can also affect how users formulate their opinions [26, 28].

In this paper, we explore whether online video adequately supports the nuanced communication and interpretation that typically happens during traditional needfinding. Our intention is not to eliminate face-to-face needfinding, but to

CSCW '15, March 14 - 18 2015, Vancouver, BC, Canada
Copyright 2014 ACM 978-1-4503-2922-4/15/03 $15.00
http://dx.doi.org/10.1145/2675133.2675174

explore the potential complementary benefits of online methods. First, we conducted a between-subjects feasibility study with 18 participants to contrast in-person needfinding with online video and text-based approaches. We learned that users generally preferred recording their feedback with video rather than typing, as it took less time and effort to verbalize their thoughts and preserved some of the richness of face-to-face interaction. We also identified several factors—such as users feeling timid and providing rambling responses—that could affect online feedback systems.

Second, these findings informed the design and development of an online system for early-stage needfinding through asynchronous video. The system utilizes Amazon's Mechanical Turk (mTurk) [1] to recruit a diverse population of users. It allows designers to tailor and launch their needfinding tasks and to analyze the responses.

Third, we conducted a case study with a professional design team who was creating a mobile reminder app. We compared the speed, sample size and outcomes of needfinding activities performed on our system versus via conventional face-to-face interviews under an equivalent budget constraint. Results show that while each individual online participant generated fewer unique considerations (40% less), collectively, the online approach revealed just as many if not slightly more unique design considerations. Further, the design team discussed the benefits of a more diverse population and how this affected their prioritization of features for a new mobile service. We conclude by discussing future research directions for online needfinding.

RELATED WORK

In this section, we review current practices of early-stage needfinding, and discuss how different media can affect the exchange of design feedback.

Practices for Early-Stage Needfinding

It is essential for designers who target a broad audience to identify the needs common to a global set of users, as well as the preferences of local markets, before settling on a final product or service [13]. Exploring alternatives during the early stages of a design project can help designers solve the right problems [16, 28] and to avoid overinvesting in a particular design direction [13, 16].

Early-stage needfinding can encompass a broad set of methods, including focus groups, field observations, surveys, or interviews [22]. For example, "speed dating" is a rapid-fire technique for exploring early-stage design alternatives where users react to a series of concept storyboard [12]. Designers conventionally conduct these activities with local, co-located participants. This allows designers to instigate a dialogue with users to understand their underlying dreams, values, and needs, [12, 33]. Face-to-face interaction allows designers to pick up on a variety of subtle emotional cues [11] and to interpret possible design opportunities in situ. For example, Tohidi et al. learned that participants worry about offending designers during face-to-face feedback sessions [28]. Designers need to be able to interpret if users favor a design alternative just because they want to appease the designer [28].

This work considers the tradeoffs of conducting early-stage needfinding online. Many of the subtleties of face-to-face communication will be lost, especially when using asynchronous communication [11, 8]. However, designers might be able to more quickly and inexpensively discover user needs by taking advantage of online participants. Online video may provide a means to increase the communication richness while still tapping into a large diverse set of users.

Indeed, video-based crowdsourcing services that collect consumer inputs via webcam such as MindSwarms [24] have gained in popularity. However, designers lack evidence that such an approach provides value for conducting early-stage needfinding activities. This paper aims to fill the gap by exploring whether online collaboration could adequately reveal users' needs, in terms of both quantity and quality, under budget similar to usability testing [25].

How Online Media can Affect Feedback Exchange

Communication medium likely influences how and what feedback providers are willing to share with designers. Hebert and Vorauer showed that critics tended to be more positive when giving specific feedback in person rather than over email [17]. Similarly, Bietz found that writers prefer to hear and see critics (vs. using instant messaging) [7]; writers interpreted text-based feedback as more negative and less credible. Literature on computer-mediated communication indicates the audio-video channel can help create a greater sense of social presence [4], and in turn, establish the trust necessary for an honest exchange of feedback [8].

While video seems valuable for online communication related to design, it may introduce challenges related to anonymity [26] and evaluation apprehension [7]. Feedback providers may be nervous about recording video [4], or prefer to remain anonymous so they have license to be more critical with their feedback.

Our research explores how online video compares with face-to-face and text-based approaches to needfinding. Specifically, we investigate whether online video adequately supports the nuanced interpretation of user reactions to early-stage concepts.

FEASIBILITY STUDY: ONLINE FEEDBACK COLLECTION

To generate insights for an online system for early-stage needfinding, we conducted a feasibility study with three research questions:

1) Do online feedback providers prefer to provide insights via text or video?
2) Will online video adequately capture the nuanced communication that typically happens during in-person needfinding?
3) What considerations affect the design of an online system dedicated to early-stage needfinding?

	Pros (# of participants)	Cons (# of participants)
Face-to-Face	Fast interaction (12) Easy to get points across (8) Easy to clarify ambiguous information (7) Non-verbal cues available (6) Personable, stronger social presence (6)	Reluctant to say more about personal experiences (5) Reluctant to say negative things about the design (5)
Video	Non-verbal cues available (6) Fast interaction (5) Personable, strong social presence (5)	Technical issues e.g. lag or poor sound quality (7) Awkward / weird to talk to a camera (4) Can only see upper torso; incomplete non-verbal cues (2) Intimidated / distracted by the self-view (3)
Text	Not intimidated by seeing each other (8) More time to think (6) Anonymous (3)	Slow interaction (7) ; Takes effort to type (5); spelling issues (3) Non-verbal cues not available (4); it's as not personable (4) Broken flow of communication (3); may forget points (3) Uncomfortable with a permanent record of their feedback (3)

Table 1. Pros and cons of the three communication methods summarized from the post-simulation interviews.

Feasibility Study: Procedure

Simulation of Needfinding Activities

To simulate a speed dating session [33], we generated storyboards using Autodesk SketchBook [2] to depict three scenarios around domestic health: home monitoring, medical records, and diet and exercise (see Figure 2 for an example). Participants reflected on their experiences, perceived needs, and knowledge of existing solutions regarding each of the design scenarios.

Using a within-subject study design, each participant experienced three ways of providing feedback on the storyboard scenarios: 1) talking to a person sitting across from them; 2) talking into a webcam and having their responses recorded; and 3) typing down the answers on Etherpad Lite, an online documentation tool [15]. We counterbalanced the assignment of scenarios and the communication methods.

Post-Simulation Interview

Upon the completion, we interviewed all participants in person and gathered their general reflections, preferences, and thoughts on comfort and ease of use, as well as pros and cons on the three different feedback methods.

Participants

We recruited 18 participants (7 female, ranging from 18-65 years old, from a diverse ethnic background) using a local participant pool and word of mouth. We also pre-screened participants to ensure that they were somewhat familiar with the three health-related scenarios, spoke English as a primary language, and had average (or better) typing skills.

Feasibility Study: Findings

Although all participants picked face-to-face as their favorite communication modality, two thirds of them preferred video over text when having to respond online. Similar to Barksdale et al.'s findings with temporally distributed teams [4], participants noted that, with video, they could present their thoughts faster, and convey complex or subtle messages via gestures and facial expressions (Table 1).

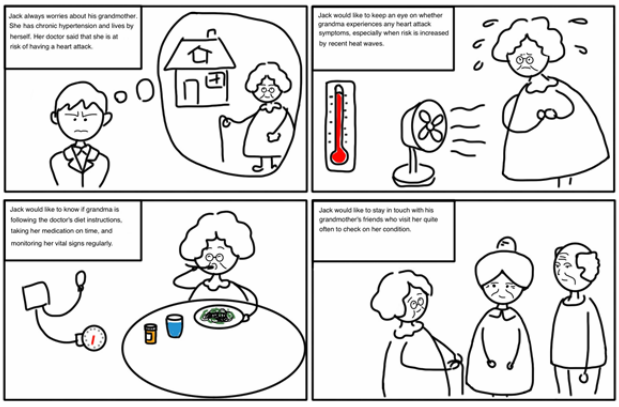

Figure 2. An example four-panel storyboard in Study 1. This story depicts the idea of monitoring health at home.

Interestingly, about 56% of the participants felt that they would more boldly talk about negative experiences or feelings if they did *not* see the immediate reaction of the other party, similar to results presented by Sheer and Chen [26]. According to one participant:

> [I would not] say something negative, probably not face-to-face. Because you want to have the courage to say something bad, you don't want to hurt the person's feelings…

This suggests that even though participants were generally more familiar with face-to-face interactions, online communication may afford a degree of "distance" between designers and users. The reduced social presence of online media compared to face-to-face interviews [8] may encourage critical feedback on designers' ideas. Among those who were not reluctant to share negative information online, participants were (1.5 times) more likely to favor video than text.

Four participants mentioned that they were nervous in front of a camera (Table 1), "*I was kind of conscious of how my hair looks in the video all the time.*" This notion of social awkwardness did not bear out when participants provided

text feedback. Three people appreciated its affordance of a greater degree of anonymity and better privacy protection:

> It (text) is definitely the most anonymous. You cannot put a name to a face regarding what they say, so they're probably freer to give their feedback, because they can't be judged since the person isn't actually there.

While text does offer anonymity, designers might suspect the legitimacy of a response if the associated personal identity were less visible, as suggested by Bietz [7].

Beyond social factors, there were several usability issues that might affect an online system for collecting early-stage design feedback. First, across the different techniques, we noticed that many feedback providers needed guidance to provide responses. Over half of participants asked for further clarification or elaboration about the storyboards during the face-to-face interviews. The quality of online feedback may degrade if participants cannot receive prompt assistance from designers [5]. Second, text-based feedback providers would often wrap up their answers in a couple of sentences, or simply put down bullet points. Third, video recordings tended to be verbose and yet relatively content-free. Participants generated almost three times more words per minute in video (75.3) than in text (26.3), but there was no significant difference between the numbers of unique points conveyed using the two media. Some participants suggested that, compared to video, the rehearsability and reprocessability of text allow them to better organize their thoughts and reduce incoherent responses, as observed in previous research [4].

In summary, early-stage feedback activities involve not only exchange of information, but also an interpretation of nuanced user reactions. Our preliminary study confirms prior research on computer-mediated communication and provides evidence that Video-based communication can potentially serve as a means to collect early-stage feedback from online users. Video technology allows designers to collect rich feedback online and to understand subtle reactions to their ideas. This approach also seems to support a degree of separation between the designers and respondents, perhaps enabling feedback providers to be more candid with their input. However, for such an approach to succeed, especially asynchronously, we need an online system that carefully structures the feedback process.

A SYSTEM FOR ONLINE NEEDFINDING WITH VIDEO

To enable large-scale participation, we designed the system to collect online video *asynchronously*, rather than through live video chat. This helps minimize the time commitment and allows data to be collected in parallel, but it provides additional constraints that may affect the quality of such an approach. Given the usability issues discovered in the feasibility study and our decision to support asynchronous data collection, we present several design considerations for online needfinding via asynchronous video.

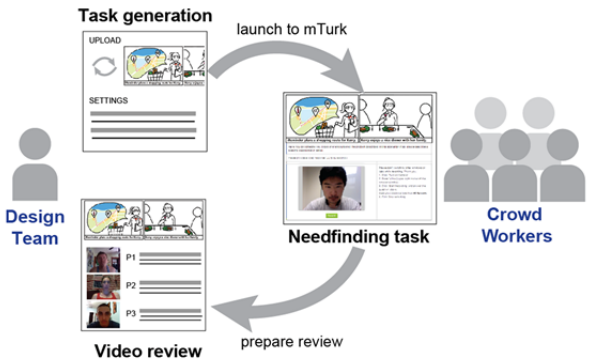

Figure 3. Overview of our online video-based system for collecting early-design feedback.

Design Considerations

Provide Examples to Model Effective Feedback

To set feedback providers' expectations, our system provides positive and negative examples of responses. These examples serve several purposes. First, they show participants the level of detail preferred. Second, examples demonstrate how participants should position themselves in front of the camera with proper lighting in the background. Third, the examples convey the impression that ordinary individuals can talk about their perception, hopefully making participants more confident about the task.

Constrain Data Collection to Yield Concise Responses

To avoid both very short and long, rambling responses we got in the feasibility study, constraints on the length of video responses should be provided. Each response should be no shorter than approximately 20 seconds and no longer than 60 seconds. The examples model the appropriate tone and length for responses. While recording a video, the system gives feedback and visual cues when providers reach the lower and upper time limits.

Allow Rehearsal to Reduce Incoherent Responses

To help respondents provide more coherent feedback, we include a simple re-record function. Likewise, the system allows feedback providers to write a short text outline before recording a video. By rehearsing their response, people do not have to organize their thoughts on the fly. It also reduces the chance of having to redo the whole video if they are not satisfied with a certain part of the response.

System Overview

The system has three parts (see Figure 3): a task generation tool for designers, a crowdsourcing component that automatically posts tasks to Amazon Mechanical Turk (mTurk) through its API [1], and a video review page for designers to view and reflect on the results. The system runs in the Google Chrome browser and was developed using HTML5, Django, and MySQL.

Task Generation Interface

Designers can set up a needfinding activity by filling out a web form (see Figure 3, upper left). Designers can upload and then describe concept sketches or storyboards and list

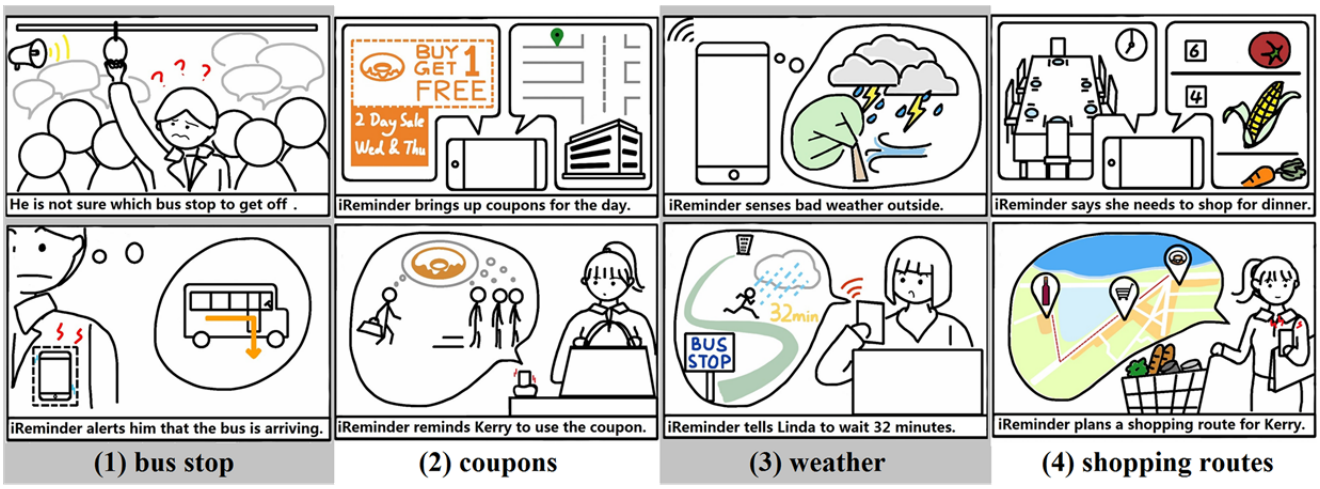

Figure 4. Sample storyboard panels for Study 2. The team explored four reminder scenarios: bus stops, coupons, weather, and shopping routes. The panels in the top row represent the problem; the bottom row depicts the potential solution.

the questions for the crowd workers. They can also specify various crowd task settings, such as the minimum and maximum length of video responses, the number of feedback providers, the amount of payment per response, the study duration, and screening questions for recruitment.

Crowd Needfinding Task Interface

After the designer configures the task, our system automatically generates the human intelligence task (HIT) for mTurk, and allows designers to monitor the progress and throughput. The system currently uses a dedicated mTurk account to post jobs, but future versions will allow designers to provide their own mTurk account details or to use other participant recruitment methods. Once the task has been accepted, workers watch a short tutorial on recording effective videos, including good and bad example of providing feedback. Workers then view the storyboard and a list of questions specified by the designer. They are prompted to briefly outline each response and then to record a short video (see Figure 3, right). The worker can preview and re-record their video if they wish.

Video Review Interface

Our system provides a video review interface for designers to analyze all user responses for each given scenario (see Figure 3, lower left). The page shows the storyboard presented to the feedback providers followed by one or multiple rows of crowd responses. Each row contains the worker's demographic information as well as the video (or text) responses to the designer's questions. Within this interface, the designer can rate the quality of each response.

CASE STUDY: ONLINE VS IN-PERSON NEEDFINDING

To evaluate our system for collecting early-stage design feedback and to understand its efficacy in comparison with conventional face-to-face methods, we conducted a case study with a professional design team working on a mobile app. We addressed the following research questions:

1) How does feedback obtained through asynchronous online videos compare to in-person interviews under a similarly constrained budget?
2) How would professional designers assess and interpret feedback from online participants, and use it to shape their design decisions?

Case Study: Method

We partnered with an industrial research institution and recruited a four-person design team in the process of conceptualizing and prototyping a new mobile application called *iReminder*. The team had been exploring ideas for a situation-aware intelligent reminder app that could leverage personal data on the phone, context information obtained by embedded sensors, and third-party mobile services.

The team wrote and illustrated four different storyboards to depict scenarios where people might get value from an intelligent reminder app (see Figure 4). The team wanted to know whether these scenarios represented typical needs and what other scenarios people may envision for iReminder.

To draw a comparison between needfinding via our online system and in-person interviews, the design team allocated $72 for each method. With this budget, designers recruited seven local users for the interviews, as advised by Nielson's rule of thumb for discount user testing [25]. This budget supports an equivalent of 10 hours of work at an hourly wage of $7.20 (the current US minimum wage). Therefore, we were able to recruit 20 mTurk workers for a half-hour session each. The case study's goal was to investigate the quality and coverage of responses each method could generate under the same budget.

Procedure for In-Person Participants

The design team sent out recruitment emails to mailing lists of local communities five days prior to the study. The team exchanged on average four emails with each local participant to schedule the time and location for the face-to-face sessions. They also spent about an hour each day to prepare

		In-person	Online crowd
Participant info	# of people	7	20
	Age range	22-29	18-41
	Ethnic groups	2	4
	Distance from design team	12 miles (avg.)	17 cities / towns in US
Ideas Collected	Total valid	36	56
	Ideas/Person	5.14	2.95
Time (avg.)	Task time	27 min	25 min
	Recruitment	5 days, 4 emails/person	10 min to post on mTurk
	Preparation	1 hour	N/A
	Transport time	103 min (round trip)	N/A
Payment (avg.)	Study/Person	$6.45 (1 hr)	$3.60 (30 min)
	Trip/Person	$3.84	N/A
	Total budget	$72.00	$72.00

Table 2. Tradeoffs in time and participant demographics for in-person versus online crowd participants.

the design studio, documents, and video recording equipment in the morning, and transfer data at the end of the day.

When participants arrived, a member of the design team greeted them and interacted with them throughout the study. Participants viewed the four storyboards one by one on an iPad, and had conversations with the designers regarding the needs and solutions related to each usage scenario for iReminder. The questions followed the same script as the online procedure, except that designers could ask for clarification or elaboration on specific points. After commenting on all storyboard scenarios, participants filled out a demographics survey and provided information on their daily practices related to the depicted scenes.

Procedure for Online Crowd Participants

The design team uploaded the same four storyboards to our system and then worked with our research team to configure the crowd task settings. The design team recruited US-only mTurk workers which allowed them to diversely sample their target population, i.e. Americans who are likely to be smartphone users. Once finalized, the system automatically posted HITs on mTurk. When mTurk participants accepted the task, they first reviewed a consent form that explains the study's purpose and asks them to explicitly consent to being video recorded and to allow their video recordings and images be used for research and academic publications. Next, participants checked their audio and video settings by recording a short clip. If the recording feature worked properly, participants went through a video tutorial with examples of good and bad design feedback.

Participants then read a short description about the concept and technology behind the intelligent reminder application, and viewed the design team's storyboard scenarios. For each scenario, participants were instructed to provide responses to two questions: "Have you (or someone you know) ever encountered the problem described in this scenario? If so, please describe a specific experience in detail." and "How effectively do you think the technology in this scenario solves the problem? Elaborate on the features of this technology that could help the situation." The system bounded the video responses so they could be no shorter than 20 seconds and at most 60 seconds.

After providing input on all four storyboards, participants filled out the same questionnaire that was used in the in-person interviews. Our system automatically logged all the data and constructed a review interface for the design team.

Procedure for the Design Team to Review Feedback

After collecting data from both in-person and online participants, our research team met individually with all four members of the design team to examine all 80 online responses (20 participants responding to all four storyboards) and 28 in-person responses (7 participants x four storyboards). For each response, we instructed the designer to extract and tally any valuable insights and to state why they liked or disliked about each response. We then asked each designer to reflect on the pros and cons of online vs. in-person design feedback collection and to discuss their overall perceptions of online needfinding and their expectations for future use regarding the number of participants, response duration, time to yield feedback, and overall costs. Each review session took between 2 and 2.5 hours. We audio recorded each review session, and later transcribed and extracted themes from these sessions.

Case Study: Results

To compare the cost effectiveness between in-person and online needfinding, we first review tradeoffs in time, location, and participant demographics. To analyze the value of each approach, we asked the design team to review p. Finally, we gathered the design team's qualitative reactions to participant responses and to the online feedback collection tool.

Tradeoffs in Recruiting Local vs. Online Participants

Table 2 lists the detailed differences in time, location, monetary costs and participant demographics between in-person participants and online crowds. The face-to-face interviews took 22 to 30 minutes (μ=27 min). Interviewees travelled 20 to 80 minutes (μ=51 minutes one way by public transportation) to the meeting location. Each interviewee received a cash payment (in local currency) equivalent to $6.45. They were also compensated for their round-trip public transportation ($2.60 ~ $6.41, μ=$3.84). Within a pre-set budget of $72, the design team recruited seven local participants. These participants ranged in age between 22 and 29 (μ=25.4) years and all had at least a college degree. Six of seven were from the same Ethnic group; only two were female.

In contrast, each mTurk worker spent 25 minutes on the task page. Approximately 15 minutes were spent on the consent form, examples, the actual feedback task, and final questionnaire. Participants spent the remaining ten minutes learning and testing the video recording feature. Each online worker received $3.60 for his or her participation, comparable to the minimum half-hour wage in the US. Under the same $72 budget, our system recruited 20 participants (14 female) from mTurk. Ages ranged between 18 and 41 (μ=26.9) years. Among the participants, thirteen were Caucasian, three were Hispanic or Latino, two were Asian, and two were African American. They all had received at least some college education.

Design Team's Reactions to In-Person vs. Online Feedback

Members of the design team individually analyzed and compared responses from the online and local participants to extract insights for their design. As shown in Table 2, on average, the design team drew out about 2.95 points of feedback per online participant (SD=1.12) via our system and 5.14 points per person from the in-person interviews (SD = 2.20), and these coalesced as thirteen unique and valuable insights (see summary in Table 3). While in-person interviews yielded more points per participant (likely due to their ability to converse with the design team without constraints [5]), these data show that online communication can provide a viable way to understand users' needs and practices. With only 20 online users, designers were already able to discover all issues identified by the lab-based study of a minimum size, i.e. full coverage. Online participants even raised several points that neither the design team or in-person interviewees had considered (Table 3). The team found online feedback to be as candid and insightful as from the interviews. The following quotes from online participants exemplify two points that were equally valued by both online and local participants:

> **Provide reliable and easy-to-understand information** - "I'd like this to be effective and work properly, but I'm a little skeptical about the app's ability to predict weather circumstances in such a small area." (P3, M, 23 yrs)

> **Be situation-aware and personalized** - "The system could suggest that I can pop into the nearest 7-11 and pick up a disposable poncho or umbrella, which could cost less than five pounds. Or think about alternative transportation, like go into an Underground or an indoor bus stop." (P11, M, 41 yrs)

Design Team's Interpretation and Use of Feedback

The design team considered feedback from both online and in-person participants to be trustworthy, even when they received conflicting opinions. The team found that feedback from both methods covered all aspects that they were particularly interested in: 1) Affirming the problematic scenario. All participants shared personal anecdotes about their life or situations similar to the given scenarios. 2) Envisioning alternative uses for the technology. Online participants proposed five possible use cases for other populations or situations, while the local users only mentioned two. 3) Suggesting how the app should behave. 4) Commenting on specific details of the technology, e.g. privacy protection and battery management. 5) Expressing skepticism or concerns, e.g. on accuracy and adaptability of the app.

Valuable points raised by respondents	O	L
Provide suggestions to users in addition to facts*+	8	3
Minimize the user input by linking to existing apps*+	8	5
Provide accurate, reliable, and interpretable info*+	7	6
Leave sufficient time for users to get prepared+	6	3
Be able to handle big, diverse data	6	3
Be situation-aware and personalized+	5	5
Allow for manual management +	4	0
Be especially informative at unfamiliar locations	4	6
Address privacy issues and reduce interference+	3	1
Be easy to learn and not to complicated	2	2
Be able to correct the user's mistakes*+	1	0
Have good power/battery management*+	1	0
Benefit & challenge for 3rd-party service providers*	1	2

Table 3. Frequency of 13 valuable points (as counted by the design team) generated from online crowd participants (O) vs. local in-person participants (L); those marked with * were initially overlooked by the design team; those marked with + were incorporated into iReminder.

These insights in turn affected how the design team interpreted and shaped the iReminder application, especially the ones that had previously been overlooked in their initial conceptualization of iReminder (marked with * in Table 3). For example, one designer commented:

> I didn't expect that information reliability would have such a big impact on people's adoption of iReminder. I guess we should inform people, maybe explicitly, our source of third-party information and confidence level. When the confidence is low, iReminder should use a softer tone and provide users with more options. We could even ask the users to provide information on the actual situation (Design team member 1).

After reviewing the feedback, the design team continued to develop a working prototype of iReminder, which evolved to account for several valuable points, especially those mentioned by disparate users (marked with + in Table 3).

Design Team's Appreciation of the Online Crowd's Diversity

Overall, the design team found online needfinding to be a viable way to broadly understand users' needs and practices. In particular, the team noted that online needfinding provided a diverse pool of participants from different cities, and neighborhoods, lifestyle choices, and ethnic backgrounds. Despite the diversity of participants, the feedback was dense and useful, likely due to the structure provided in the tool (i.e., examples, time limits, rehearsals, etc.).

When evaluating ideas through in-person interviews, the design team inevitably taps into local participants, people likely to share similar experiences, lifestyle, and culture. As a result, local participant populations may only provide a glimpse of the range of possible reactions to an idea. For example, the design team's office is located in a densely populated metropolis with a public transportation system that effectively shelters travelers from chaotic weather. It is common for the residents to travel by bus and subway, shop

at specific stores/malls/supermarkets within walking distance from their home, and dine out regularly near work place or home instead of cooking. Therefore, most local participants felt strongly that iReminder would be mostly useful for emergencies or when traveling to unfamiliar places. Five out of the seven in-person interviewees commented that they would use iReminder to avoid long queues in bus/subway stations and restaurants, rather than as a tool to escape the rain. By also recruiting *online* participants, the design team gained novel perspectives:

> We realized how different the transportation needs are for people in a metropolitan area versus people living in a rural area where driving is the primary means of getting around.

Our system enabled the design team to collect opinions from people with different transportation systems, public facilities, and social norms. Although the design team tried to encourage diversity in the local interviews by recruiting people from different cultural or professional backgrounds, the monetary costs and logistical issues made this difficult.

Design Team's Comments about the System

The design team felt that video gave access to the feedback providers' emotional reactions and engagement in their ideas. Video allowed the design team to pick up on subtle contextual information about the participants—information that could never be captured using text alone. Designers felt that they could make better personal connection to the feedback provider and thus value their responses as more truthful. For example, the team commented on participants' clothing and background décor. They noticed that one participant had many American flags and speculated about her nationalism. They observed one participant owned a cat, because they could see its tail.

Designers picked up on nonverbal cues as an indication of whether people resonate with a given scenario; these cues also lend credibility to the feedback [7]:

> (With video) you get to see their, like, reaction on their face, and also you can hear their voices, and you can see and hear their hesitation or any change of emotions.

While these are minor details, the video data provided a viable and cost-effective way to gain a deeper understanding of their potential audience.

The design team agreed they would use our system on future projects to collect early-stage design insights. However, they also had a number of observations about how to improve the system and crowd needfinding in general. The design team actually wanted more responses than provided by the 20 online participants recruited for this study. They agreed it would be ideal to get feedback from 50 to 100 people and to be able to produce this feedback within two weeks. Our system can handle this demand and may be able to gather feedback in a fraction of this time using a retainer model or other real-time crowd recruitment methods [6].

The design team also felt the video responses should be longer, as one team member stated:

> One-minute videos seem fine from what we just saw. But I feel that we should not set an upper limit. Participants should be able to talk as long as they want.

Although current system limits the recording time to avoid rambling responses based on our results from the feasibility study, we could remove this upper limit (or allow the designer to set it) to accommodate more information. Further, the design team expressed desire for follow-up conversations with particularly insightful respondents:

> I'd love to hear more about the elderly story" or off topic "It is a pity that this person didn't say much as he didn't realize that we have an intelligent engine in the backend.

The designers wanted more interaction with participants, which is something we hope to address in future work. In terms of willingness to pay for such a service, team members said they may pay between $1 and $3 per response, "*the price of a cup of coffee.*" But, they would be willing to pay a bonus up to $10 for exceedingly valuable feedback. "*I am willing to pay more… and I can wait for a longer time as long as there are more interesting responses rolling in.*"

Case Study: Discussion

This research demonstrates how online communication can work *in parallel* with in-person interaction for designers to quickly cast a wide net to capture new insights and discover previously overlooked issues in early-stage design [10]. While designers sometimes felt less sense of presence from online users compared to talking face-to-face and online users generally yielded fewer ideas per person, the crowd as a whole generated more diverse insights. We found that twenty online participants can sufficiently identify issues discovered by seven local interviewees. This approach is also time-and-cost-effective because professional design teams do not need to pay for transportation costs, spend days recruiting participants, or setup a space for interviews.

The study shows that crowd-sourced asynchronous video communication is a viable method for designers to iterate both on their designs and their target audience who are not cultivated for design research. This work provides evidence that workers on mTurk can and will record useful self-reflections on video. For the most part, workers followed our examples to effectively position themselves in front of the camera and to outline their thoughts before capturing the video. Workers re-recorded their videos a total of 28 times out of 100 showing the utility of the rehearsal feature. With properly designed mechanisms, asynchronous online video can successfully support the rich interactions required for early-stage needfinding.

Overall, the design team was impressed by the personal nature of mTurk workers' stories and wanted to learn more. Future versions of our system would allow designers to contact selected workers for follow-up interviews. We also plan to make the system more robust by allowing workers to save their progress and providing features like crowd-enabled speech-to-text transcription to give designers a way to preview the content of videos by reading text.

FUTURE RESEARCH ON ONLINE DESIGN PRACTICES

Our online feedback system can be used in any type of design: product, service, interaction, etc. In future work, storyboards can be replaced by other digital design artifacts such as videos and slideshows, further strengthening the generalizability of this approach. Based on the results and comments received from the two studies, we present some future directions for conducting general design research activities with remote users [10]:

Create effective design artifacts and prompts. This can help feedback providers understand the task, whether it is about a concept, a scenario, or an interaction, and reduce the noise in the feedback. We observed and learned from the post-study interviews in both studies that feedback providers might misinterpret the scenario or miss some of the key messages when they had trouble following the storyboard. This could happen in usability testing and other design activities as well. Therefore, it is critical for the design artifact and prompts that participants interact with to stand-alone, to be easy to understand, and to follow known guidelines (e.g. [29, 30]). Designers can pilot their artifacts and prompts with several participants to test their readability and interpretability before presenting to a larger crowd.

Support more recruitment channels and demographic filters for pre-screening and post-analysis. Designers should identify their target audience and choose platforms or services — crowdsourcing sites, online forums, social networking sites, etc. — that are most likely to provide access to these users. Once people sign up, pre-screening could help ensure quality feedback. Through simple survey methods, designers could capture participants' demographic information, cultural background, and familiarity with the scenarios or design. In addition, when choosing the communication modality, designers should take into consideration target users' literacy and language skills, accessibility to technology (e.g. webcam and decent Internet speed for video chat), and technology fluency (e.g. typing speed). These data could allow designers to pre-screen participants to fit certain criteria or they can be used during data analysis to examine the differences between disparate user populations (e.g., urban vs. rural). As the design activity proceeds, such an approach would help designers flexibly target specific or new participant pools.

Enable information exchange through follow-up strategies. As the case study revealed, designers often want to dig deeper and interact with participants. This is a clear benefit of face-to-face interviews. Our system may support better feedback communication by allowing designers to send follow-up questions to particular users. Going further, online design communication can incorporate both synchronous and asynchronous channels. To take advantage of the scale and relative ease of crowdsourcing, we envision this as a two-stage process. First, the system recruits participants to leave short video responses about designers' concepts (this is currently implemented in the system). Second, if the designer wants to follow up with particular users, the system would help schedule a live video chat between the two parties (future work), much like an in-person interview.

Increase designer presence to promote participant engagement. Designers' degree of social presence, defined as their salience in the mediated environment, may in return affect participants' perception, engagement, trust, and level of satisfaction around the design activity [27]. Online communication systems could further emulate face-to-face interactions by increasing designers' social presence, such as displaying a personable identity or provide voice feedback/appreciation when participants finish the task.

Protect participants' privacy and designers' intellectual property. Privacy features would make both parties more comfortable exchanging ideas online. In the feasibility study, many participants worried that their chat history or video footage would be part of a permanent record. The system should take measures to protect anonymity—for example, by obfuscating the video data and hiding personally identifying information—especially when performing design activities on sensitive topics. Informed consent should be obtained from crowd participants if video data will be used in communication with other stakeholders (our studies followed proper human-subjects protocols). Likewise, the design team in the case study raised the point that presenting novel ideas to a potentially anonymous online crowd raises the risk of leaking intellectual property. The design team proposed that in certain circumstances, participants would need to sign a Non-Disclosure Agreement.

CONCLUSION

This paper showed that, under a similar budget constraint, online crowds can provide early-stage design feedback through asynchronous video with a quality and coverage comparable to face-to-face methods. Through a feasibility study, we first explored the effect of media on feedback providers' experience, comparing face-to-face interviews, online video communication, and text-based chat. Lessons from this feasibility study guided our development of a system for designers to collect asynchronous video feedback from online crowds. We studied its potential impact through a case study with a professional design team and compared the efficacy of online needfinding versus conventional in-person interviews. While the in-person interviews produced more insights per participant, online video responses were cheaper and easier to collect and allowed designers to yield more overall insights from a more diverse pool of people. The design team reacted positively to online video communication, and indicated their desire to use the system for future design projects; they also provided suggestions on improving online feedback collection.

ACKNOWLEDGMENTS

Funded by National Science Foundation grants #1217096 and #1208382, CI Fellows grant #1019343, and REU award #1340291. We thank the design team and all participants.

REFERENCES

1. Amazon's Mechanical Turk. (2014). https://mturk.com/.

2. Autodesk SketchBook. (2012). http://m.autodesk.com/sketchbook/.
3. Andreasen, M.S., Nielsen, H.V., Schrøder, S.O., and Stage, J. 2007. What happened to remote usability testing? An empirical study of three methods. In *Proc. of CHI'07*. ACM (2007), 1405-1414.
4. Barksdale, J., Inkpen, K., Czerwinski, M., Hoff, A., Johns, P., Roseway, A., and Venolia, G. Video threads: asynchronous video sharing for temporally distributed teams. In *Proc. CSCW'12*. (2012) 1101-1104.
5. Bastien, C. Usability testing: a review of some methodological and technical aspects of the method. *International Journal of Medical Informatics*, (2010). 79(4):18-23.
6. Bernstein, M.S., Karger, D.R., Miller, R.C., and Brandt, J. Analytic methods for optimizing realtime crowdsourcing. In *Proc. Collective Intelligence*. (2012).
7. Bietz, M.J. Effects of communication media on the interpretation of critical feedback. In *Proc. CSCW*'08, ACM Press, (2008). 467-476.
8. Bos, N., Olson, J., Gergle, D., Olson, G., and Wright, Z Effects of four computer-mediated communications channels on trust development. In *Proc. CHI'02*. ACM Press, (2002). 135-140.
9. Buxton, B. *Sketching User Experiences*. Morgan Kaufmann Publishers, (2007).
10. Cross, N. Design cognition: results from protocol and other empirical studies of design activity. In: Eastman, C.; Newstatter, W. and McCracken, M. eds. *Design knowing and learning: cognition in design education*. Oxford, UK: Elsevier, (2001). 79–103.
11. Daft, R. L. and Lengel, R.H. Organizational information requirements, media richness and structural design. *Management Science*, (1986). 32 (5):554-571.
12. Davidoff, S., Lee, M.K., Dey, A.K. & Zimmerman, J. Rapidly Exploring Application Design Through Speed Dating. In *Proc. UbiComp2007,* (2007). 429-446.
13. Dow, S., Fortuna, J., Schwartz, D., Altringer, B., Schwartz, D.L., Klemmer, S.R. Prototyping Dynamics: Sharing Multiple Designs Improves Exploration, Group Rapport, and Results. In *Proc. CHI'11*. (2011).
14. Dow, S., Gerber, E., and Wong, A. A pilot study of using crowds in the classroom. In *Proc. CHI'13.* ACM Press. (2013). 227-237.
15. Etherpad Lite (2014). http://etherpad.org.
16. Gaver, B. and Martin, H. Alternatives: Exploring Information Appliances through Conceptual Design Proposals. In *Proc. CHI'00.* ACM Press. (2000).
17. Hebert, B. G., and Vorauer, J. D. Seeing through the screen: Is evaluative feedback communicated more effectively in face-to-face or computer-mediated exchanges? *Comp.in Human Behavior*, 19(2003): 25-38.
18. Hui, J., Gerber, E., Dow, S. Crowd-Based Design Activities: Helping Students Connect with Users Online. In *Proc. DIS'14*. (2014).
19. Kittur, A., Chi, E., and Suh, B., Crowdsourcing for usability: Using micro-task markets for rapid, remote, and low-cost user measurements. In *Proc. CHI'08*. ACM Press, (2008).
20. Kohavi, R., Henne, R. and Sommerfield, D. Practical guide to controlled experiments on the web: listen to your customers not to the hippo. In *Proc. SIGKDD'07*, (2007).
21. Liikkanen L., Kuikkaniemi K., Lievonen P. and Ojala P. (2011) Next Step in Electronic Brainstorming: Collaborative Creativity with the Web. In *Proc. CHI2011.* ACM Press. (2011).
22. Martin, B., Hanington, B. *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Rockport (2012).
23. Martin, R., Shamari, M.A., Seliaman, M.E., and Mayhew, P. Remote Asynchronous Testing: A Cost-Effective Alternative for Website Usability Evaluation. *International Journal of Computer and Information Technology*. (2014). 3(1): 99-104.
24. MindSwarms. (2014). https://www.mindswarms.com/.
25. Nielsen, Jakob, and Landauer, Thomas K. A mathematical model of the finding of usability problems. In *Proc. INTERCHI'93*. ACM Press. (1993). 206-213.
26. Sheer, V. and Chen, L. Improving Media Richness Theory : A Study of Interaction Goals, Message Valence, and Task Complexity in Manager-Subordinate Communication. *Management Communication Quarterly*. (2004). 18 (76): 76–93.
27. Short,J.,Williams,E.&Christie, B. *The social psychology of telecommunications*. London: Wiley. (1976).
28. Tohidi, M., Buxton, W., Baecker, R., and Sellen, A. Getting the right design and the design right. In *Proc. CHI2006*. ACM Press. (2006). 1243–1252.
29. Truong, K. N., Hayes, G. R. and Abowd, G. D. Storyboarding: An Empirical Determination of Best Practices and Effective Guidelines. In Proc. *Proceedings of the 6th conference on Designing Interactive systems*. (2006), 12-21.
30. van der Lelie, C. The Value of Storyboards in the Product Design Process. *Personal and Ubiquitous Computing*. 10 (2). (2006). 159-162.
31. Xu, A. and Bailey, B. A crowdsourcing model for receiving design critique. In *Proc. CHI'11*. (2011).
32. Yu, L. and Nickerson, J.V. Cooks or Cobblers? Crowd Creativity through Combination.In *Proc.CHI'11.*(2011).
33. Zimmerman, J. Designing for the self: making products that help people become the person they desire to be. In *Proc. CHI'09*. ACM Press, (2009). 395-404.