

Evaluating Paper Prototypes on the Street

David G. Hendry
Information
School
University of
Washington
Seattle, WA 98195

Sara Mackenzie
Department of
Family Medicine
University of
Washington
Seattle, WA 98195

Ann Kurth
Biobehavioral
Nursing & Health
Systems
University of
Washington
Seattle, WA 98195

Freya Spielberg
Department of
Family Medicine
University of
Washington
Seattle, WA 98195

Jim Larkin
Resources Online
3524 Bagley
Avenue North
Seattle, WA 98103

{dhendry, saramack, akurth, freya}@u.washington.edu jim@ronline.com

ABSTRACT

The evaluation of paper prototypes is normally conducted in controlled settings such as a usability lab. This paper, in contrast, reports on a study where evaluations of a paper prototype were performed on the street with young adults. We discuss the merits of this approach and how it impacted the design process. A key finding is that the street location can enfranchise people who may otherwise be under-represented in design. We conclude that evaluating paper prototypes in public, street settings is feasible and informative.

Author Keywords

Paper Prototyping, Design Research, User Research, Usability, Health Informatics, Sexual Health

ACM Classification Keywords

H.5.2 [User Interfaces]: Evaluation/methodology; J.3 LIFE AND MEDICAL SCIENCES

INTRODUCTION

Evaluations of paper prototypes are typically conducted in a work setting or usability lab where participants perform tasks and their performance is observed [5]. This method has been shown to be extremely effective for finding usability problems early, or indeed late, in the design process. At the same time, paper prototypes can be created and revised cheaply and, through joint observations of participants, evaluations create an informative and socially rich setting for interdisciplinary design. With low cost and high impact, paper prototyping is an effective method for the iterative development.

In this paper we report on evaluating a paper prototype on the street, a setting quite unlike an office or usability lab. First, we introduce the aims and development process for this project. Then, we present the process used to evaluate

the paper prototype. Finally, we reflect upon the merits of this approach.

DEVELOPMENT PROCESS

Design Aim: Behavioral Change

CARE¹ is a Tablet PC application for learning about sexual health. It addresses a profound human and social goal: To help people to reduce behaviors that put them at risk of becoming infected with sexually transmitted infections (STIs), including HIV. STIs disproportionately affect groups that are often disenfranchised, such as young men and women of color or of low economic standing [1].

CARE is currently undergoing field trials and will eventually be deployed in clinical and public health settings in rural and urban areas, walk-in health clinics, mobile HIV testing facilities, and so on. The physical and social factors embedded in these settings will lead to great differences in how people encounter and use CARE. In some settings CARE will complement face-to-face sexual health counseling; in others, due to limited resources or lack of trained counselors, it will serve largely as a substitute for counseling and will be used as a personal tool for behavioral change. CARE can be deployed in various stand-alone and networked configurations.

From this vision come many specific objectives. Only one, however, is examined in this paper: *How should the interaction design for CARE be structured for effective communication?* While a rich science base on sexual health and behavioral change can inform solutions to this problem, this literature is complex [1]. Questions about the interplay of biological, psychological, social, developmental, and economic forces that lead people to engage in high risk behaviors are incompletely understood. Nevertheless, it is known that brief counseling in the clinical setting can lead to behavioral changes which, in turn, reduce the risk of contracting STIs [3,4].

¹ CARE stands for **C**omputer-**A**ssessment & **R**isk reduction Education for HIV/**S**exually Transmitted Infections.

Design Principles

While equipping people with information about STI risks and safer practices, including abstinence, is one piece of any solution, the more fundamental goal is behavioral change. One way to frame the problem is by *technology that persuades* [2]; that is, by human-computer dialogs that lead to changes in behaviors, attitudes, and motivations. Three principles from this orientation inform the design:

1. *Tunneling*. CARE presents a linear process: Options for browsing are minimized and even backtracking along the path is discouraged.
2. *Tailoring*. CARE tailors the presentation of material to what people have said about themselves. The further down the tunnel one goes, the more familiar it seems.
3. *Credibility*. CARE seeks to build credibility by anticipating sources of concern that many people express about health institutions, such as confidentiality.

CARE presents people with a linear path, divided into four phases. The first, *Welcome*, familiarizes people with the operation of CARE and allows people to register anonymously for repeated use. Next, in *My Risks*, people answer a branching sequence of questions about their risk factors, including age, use of alcohol and drugs, and sexual behaviors. Each question is determined by the responses already given and word choice is carefully tailored to, for example, gender and sexual orientation. Then, in *Thinking it Through*, people are first informed of their top 3–5 risks, which are estimated through a weighing scheme using epidemiological and clinical data. Next, risk reduction practices are modeled using brief videos that address their specific barriers to sexual safety (e.g., how to talk to a sexual partner about abstinence, condom use, SDI/HIV testing, and so on). Finally, in *My Plan*, people select a single behavior change goal, and identify a list of specific action items for achieving the goal. At the end of the session, people receive a printed report as a reminder of their goal and action items. This report can also be used to begin a conversation with a counselor or it can be kept private. A CARE session takes approximately 30 minutes.

EVALUATING THE PAPER PROTOTYPE

Participants and Settings

Two key constituencies for CARE are young adults who are homeless, a term that denotes various states of habitation, and men who have sex with men. Both populations are at high risk for STI/HIV. We used purposive sampling and sought individuals in the settings where they were most likely found. The settings were:

1. On the street in an urban environment frequented by the homeless. Here, we intercepted four people who appeared to be men, over 18 years old (one participant was identified as a woman after the evaluation began).

2. At a mobile HIV testing site, located in a large camper van. Here, outreach workers solicited people from bars and eateries late at night. Users of the van site tend to be men who have sex with men. At this site, three men and one transgender person participated.

In both cases we took the paper prototype to environments that were considered safe by these participants rather than asking them to come to a usability lab or clinic. Two of us, Mackenzie (a Family Practitioner specializing in youth sexual health) and Hendry (a researcher in HCI), conducted these evaluations. Participants were compensated \$25 US.

At the street setting, we intercepted participants and sought interest by describing the study, the expected time commitment, and the \$25.00 US honorarium. To help with this, a young woman who was pan handling volunteered to inform her acquaintances of the study. She, opportunistically, became a (paid) intermediary in this project. We obtained informed consent verbally.

Paper Prototype and Process

The prototype consisted of a cardboard box, about the size of a Tablet PC, and a sequence of 20 screens which were high-fidelity renderings, printed on a color printer. Using laminated cards and transparencies, we simulated the key input/output dialogs and tailored the interface elements that persisted across multiple screens (e.g., customized names). To simulate the stylus, participants used a ballpoint pen.

The evaluation was driven by a scenario and this task: *Assume you are Joe and that you are waiting for an appointment at a sexual health clinic to begin. Now, spend the next 30 minutes or so using CARE.* The key facts about Joe were outlined on a sheet of paper and as participants progressed through CARE, they found facts in the scenario that were needed to answer questions, select videos, and pursue goals. Participants were prompted to verbalize and were given a brief practice session in thinking aloud. The session ended with a structured interview and a reflective conversation on any critical incidents encountered.

Process

One of the moderators both facilitated the evaluation and acted as ‘the computer’, the other moderator took notes. In the street setting, simulating the computer was tricky because participants held the prototype in their laps and there was no ready place to keep the paper dialog pieces. Pieces were placed on the ground and in coat and knapsack pockets—not easy on a windy day. To ensure participant anonymity, neither audio recordings nor photographs were taken. After each session, we reviewed the notes and added key observations and issues to a growing list. This material was then used to write a short, informal report where observations, issues, and sketches of design remedies were organized by screen.

Procedural Factor	Street	Lab
Target use setting	No	No
Participant recruiting	Opportunistic/ Intercept	Planned/ Scheduled
Moderator control	Limited	High
Operational convenience	Low/Variable	High/Stable
Team observation	Not possible	Direct
Data capture options	Limited	Many

Table 1. Street versus lab: Six key factors.

DISCUSSION

The procedural factors weighing against the street were considerable (see Table 1) and gave the team pause. Even more, we are unaware of any usability studies that have been conducted with homeless young adults in street settings. Would such a study be practical? Would it generate useful information? The conviction that it was important to involve these participants swayed the team's decision and, despite the procedural barriers, the study proved remarkably successful. First, the study uncovered many usability issues that led to significant design changes. Second, it improved the team's appreciation for its target users. Next, we present reflections and lessons learned.

Enfranchisement

The team recognized that important users of CARE were less likely to be insured or have a primary care professional, and were more likely to be suspicious of professional settings. How these characteristics would influence the user interface was unclear and of particular interest to the team. Responding to this general concern, the team decided that it was critical to enfranchise people who, due to these very characteristics, would be difficult to recruit to a usability lab or medical office.

Once we obtained informed consent verbally, we invited each participant to join us at a nearby coffee shop. To our surprise, each participant declined; instead, they confidently guided us to a nearby location for semi-private conversations. We conducted evaluations in an alley, on the steps of a tattoo parlor, by a bus stop, and on a public basketball court. Remarkably, in retrospect, we sat on the ground for each of these four evaluations. That participants were unwilling to go half a block to our perceived "more appropriate setting" supports the view that recruiting these participants to lab or medical office would be difficult.

From the study came telling, empathic stories that helped the team better understand the design. For example, one participant, who claimed that he had outstanding warrants for his arrest in another state, challenged the claim that the registration process was anonymous and offered another solution. Another was unsure of his father's first name, a fact that was needed by the registration process. Another

participant ignored the registration questions, and typed out a phrase. These details from specific people raised key issues that we felt would generalize to other target users. The reactions forced the development team to reconsider issues of credibility that should be addressed by CARE.

Lesson: Performing usability evaluations on the street is feasible and can enfranchise under-represented people.

Empowerment

The simple act of selecting a study location was extremely important for it signaled to participants that they would have a significant degree of control. Participants gained additional power for shaping the situation by responding to various interruptions. These included a pedestrian who asked directions to a local sex shop, a worker's jackhammer, and a friend of a participant who happened to see us and wanted to know what was happening. A slowly moving police car caused us pause during one session but the participant, obviously familiar with the neighborhood, assured us that the officer was unlikely to stop. Thus, the street induced a duality of control: The participants guided the moderators in the street setting and the moderators guided the participants through the evaluation.

We believe that the participants were empowered by this control and these young people were obviously cognizant of the issues surrounding sexual health and were genuinely enthusiastic to give input. Indeed, they appeared proud to contribute. We believe that this empowerment and enfranchisement, together with a paper prototype of ambiguous status, led to particularly critical interpretations and freer improvisations. This, in turn, led to greater insight into the design.

Lesson: Consider how usability testing can be shaped to impart a duality of control between participants and moderators. The characteristics of the street naturally enabled this sharing of power. In a medical office or usability lab it would be more difficult to achieve.

Identifying Usability Risks and Issues

As with lab evaluations, the moderators were able to elicit feedback on specific usability issues. For example, participants readily discovered the navigation controls and used them appropriately, and participants reacted positively to the overall tone and look and feel of the prototype as evidenced by spontaneous smiles and positive verbalizations. Problematic interactions were also readily observed; for example, participants had difficulty understanding some of the questions as evidenced by lengthy pauses, requests for clarification, and other breaks in the flow. Thus, both positive and negative features of the prototype became readily apparent.

Lesson: While it is more difficult to manage the evaluation process on the street, usability issues can be readily identified. This was not obvious to us prior to this study.

Interactive Ambiguity and Improvisation

Fortuitously, the prototype inadvertently contained several sources of ambiguity that caused participants, and the moderators, to improvise in interesting ways. One source is *clickability*, which refers to the degree of certainty about whether an element on a user interface can be clicked. It is signaled by such verbalizations as “Can I click on that?” or when some non-clickable element is clicked. For example, when we saw 9 of 10 people click on something that was *not* intended to be clicked, we learned that people wanted to click on it and we asked “Why is that?” Following up this discovery with team discussion led to proposals for a restructuring of this part of the interaction and improved signaling about the intended roles for content units.

A second source of ambiguity is *progress feedback*, which refers to how people are informed of their progress. It is signaled when people verbalize uncertainty for what they expect to happen next or when they verbalize a confident but erroneous expectation. For example, one participant verbalized a detailed list of safe sexual practices as if he were anticipating and rehearsing what was to come. This was unexpectedly triggered by a screen. This observation, and others like it, led to a reconsideration of how transitions between the four phases of CARE are signaled.

The third source of ambiguity is *constraint feedback*, which refers to limited feedback provided by dialog controls. It is signaled by such verbalizations as “Can I do X and Y?” when confronted with dialog controls. For example, we saw a majority of people click on two or more elements in a radio button control or improvise much richer forms of navigation than intended. We saw this not as a limitation of the mockup but as an opportunity to ask “People should be narrowing their choices—why do they want to explore so much?” This observation led to significant changes to the underlying structure of this part of CARE.

On reflection, we think that the most important insights came from the following pattern: UI elements of uncertain meaning can lead participants to experience ambiguity, which, in turn, can lead to improvisation, which, in the end, provides insights into how people are conceptualizing the artifact. Importantly, ambiguity can be injected equally well into paper-based and operational mock-ups. In sum, this analysis argues for substituting the high- versus low-fidelity characteristic of prototypes (e.g., renderings and operational systems versus sketches) with a typology for characterizing forms of ambiguity that can be deliberately exploited when creating and evaluating prototypes. The provisional vocabulary above is a small step in this direction.

Lesson: Consider how ambiguity can be deliberately incorporated into prototypes and how moderators can exploit it through improvisation during evaluations. We do not think the above issues would have been as easily identified in evaluations of a high-fidelity, operational system and tightly scripted evaluation process.

CONCLUSION

What status should be assigned to these evaluations? They are not ethnographies, nor even participant observations, because the intervention was intrusive and short. They are not generative, participatory design exercises because the design was largely established. They are not usability evaluations, as normally considered, because of the lack of control. They lie, in fact, at an interesting middle ground of enfranchisement, power sharing, and interactive ambiguity. We believe this is a ground rich in insights.

This work shows that the benefits of conducting usability evaluations of paper prototypes in transitional settings can outweigh the downsides. First, we conclude that it is entirely feasible to conduct evaluations in the street and that critical information about usability can be discovered. Second, we conclude that the street can be used to empower and enfranchise under-represented groups in design. The value of engaging such groups should not be underestimated, even if normative approaches to usability must be adjusted. Finally, the street plus an ambiguous mock-up lead to very interesting demands on participants and moderators for constructing meaning, which, in turn, leads to insight. Creating these demands in a usability lab is difficult. To improve practice, further studies should compare the results obtained from evaluations on the street with results obtained from a more controlled setting.

ACKNOWLEDGMENTS

This work would not have been possible without the efforts of Charles F. Wilson at the People of Color Against AIDS Network, Peter Frazier, Luke Celt, and the entire development team at Resources Online. Our sincere thanks go to the participants who must remain anonymous. Thanks, too, to the CHI'2005 reviewers.

REFERENCES

1. Eng, T. R. and Bulter, W. T. (Eds.). *Epidemic: confronting Sexually transmitted diseases*. National Academy of Press. Washington D.C. 1997.
2. Fogg, B.J. *Persuasive Technology: Using Computers to Change What We Think and Do*, New York: Morgan Kaufmann, 2003.
3. Kamb M. L., Fishbein M., Douglas J. M. Jr. et al. Efficacy of risk-reduction counseling to prevent human immunodeficiency virus and sexually transmitted diseases: a randomized controlled trial. Project RESPECT Study Group. *Journal of the American Medical Association* 280, 13 (1998), 1161-1167.
4. O'Donnell CR, O'Donnell L, San Doval A, Duran R, Labes K. Reductions in STD infections subsequent to an STD clinic visit. Using video-based patient education to supplement provider interactions. *Sex Transm Dis.*, 25, 3 (1998), 161-8.
5. Snyder, C., *Paper Prototyping: The Fast and Easy way to Design and Refine User Interfaces*. New York: Morgan Kaufmann, 2003.