Field versus Laboratory Usability Testing: a First Comparison

Aydin Oztoprak, Cigdem Erbug

Department of Industrial Design - METU/BILTIR/UTEST Product Usability Unit
Middle East Technical University, Faculty of Architecture
Inonu Bulvari, 06531 Ankara, Turkey

Tel: +90 312 210 4219 – Fax: +90 312 210 1251

oztoprak@metu.edu.tr - erbug@metu.edu.tr

Abstract

Whilst testing the usability of a product in a usability laboratory, usability researchers generally try to extract usability information from a representation of 'context in use'. However, there might be cases where it is difficult to simulate the real-world usage due to lack of contextual information (e.g. scarcity of information on low-socio-economic status users' living conditions, etc.), or physical limitations (e.g. difficulty in representing the actual physical environment, etc.). This set of facts requires testing in natural settings. Discussion is focused on the relative advantages and disadvantages of each setting, along with the technical infrastructure.

Gaining insight and experience in remote testing rather than searching for statistical significance was the major goal of this study. Even though this was an initial inquiry of an exploratory study, it was found that usability testing and evaluation of a product, with users' own tasks and goals, in actual use context, reveals implicit usability problems in the interface.

Introduction

In recent years, remote usability testing has been widely utilized in testing of software interfaces as an alternative to laboratory testing. Advantages are that contextual usability information is gathered, costs are lower than laboratory tests, and higher quality more relevant usability data are collected using a more representative sample of users (Hartson & Castillo, 1998, Brush, Ames & Davis, 2003, Ames, 2003, Tullis et al. 2002, Dray & Siegel, 2004, Hartson et al. 1996, Rowley, 1994, Gough & Phillips, 2003, Kantner, Sova & Rosenbaum, 2003, Petersen, Madsen & Kjaer, 2002, Wichansky, 2000).

Moreover, remote usability testing gives more realistic test setup as the participants remain in their normal setting (Brush, Ames & Davis, 2004) and helps usability practitioners in overcoming the 'homogeneous subject pool' problem emerged at many institutions by running tests with the same participants over and over (Hartson et al. 1996).

Even though laboratory testing is widely and effectively utilized in the evaluation of software interfaces, it has some limitations. Laboratory based usability studies capture a snapshot of the use in a simulated use environment. Simulating the use setting is very hard, time consuming, expensive and sometimes impossible to attain. In the laboratory, users are isolated from contextual factors (often distractions), such as interactions with other users and products which draw their attention away from the product being used. On the other hand, new electronic devices are used in a context, in which multitasking is a key factor. These devices are used while walking, talking, running or even driving. According to Thomas and Macredie (2002) these daily activities, 'distractions', should be at the centre of understanding and designing of such electronic appliances (Thomas & Macredie, 2002). Another challenge of laboratory based usability is the testing of leisure and entertainment products. For such products evaluating use experience developed over time instead of snapshot of use provides more reliable information (Petersen, Madsen and Kjaer, 2002). Remote testing emerges as an alternative to observe products being used in their natural environment with real users.

Problem Statement

Although the literature on remote usability testing activities has been growing in recent years, (Hartson & Castillo, 1998, Kantner, Sova & Rosenbaum, 2003, Brush, Ames & Davis 2004, Rowley, 1994, Gough, Phillips, 2003, Dray, Siegel, 2004, Petersen, Madsen & Kjaer 2002) most of these publications focus on the testing of software interfaces. Testing of consumer products has not attracted much attention.

Remote product testing often requires the evaluators and/or research associates to travel to the field and utilize portable usability laboratories or hire local usability laboratories at the remote site. Remote software testing however, can be achieved by various tools such as internet questionnaires, surveys,

automated usage tracking, desktop or application sharing software, desktop or video conferencing, and/or live or collaborative remote evaluation. These tools enable researchers to conduct test sessions without physically being present at the location of the subject.

To compare remote product usability testing with traditional laboratory based product testing, an experiment was conducted in both settings. The comparison was based on the number and severity of usability problems found and participants' and researchers' experience of the setting.

Testing Process

The present study is a pilot study, rather than conducting a thorough usability test evaluating all aspects of the telephone set. Goal was to gain insight and experience on remote testing of a product; thus emphasis was given on the strengths and weaknesses of the test method utilized.

The Product: A console type operator telephone set (Figure 1), with a four row LCD screen designed for professional office-use was tested. The telephone set works as a client of a central unit, consequently it only works with a specific type, brand and model of a central unit. The telephone set has alphanumeric keys (0-9 * #), shortcut keys for menu, phonebook, flash, park, transfer and redial, special function keys (16 programmable keys with double function), volume control (+,-) and hands-free keys. Four keys are reserved for navigation in the menu, two keys for up and down and two keys for the functions displayed on the screen (Accept, clear, redial, hide, etc). Feedback is provided on the four LCD rows on the screen and by sound.

Figure 1, console type operator telephone set, which was tested during the sessions

Participants: To compare remote and local settings equal number of users, five for each, was recruited as volunteers. The experimenter obtained the addresses of the institutions which use the proper central telephone units as the telephone set only works with specific types of central telephone units. The experimenter than contacted the institutions and recruited the participants. Two of the participants were telephone operators of two local radio stations, another two were architects working at a projects office and also responsible from answering phone calls, the other participants was a secretary.

There were also five users for the laboratory test. Four of them were experienced secretaries and one was a student project assistant responsible for answering the telephone.

None of the participants had previous experience with the telephone set.

Equipment and Materials: Local usability tests were carried out in a fully equipped usability laboratory (www.utest.metu.edu.tr/useiteng). Test sessions were videotaped by two cameras simultaneously, one targeted to the telephone set and the other to the face of the subject. The two video streams were than superimposed for further reviewing and analysis. Sound was also recorded during laboratory tests. Test subjects were alone in the room and the experimenter observed them through a one-way mirror from the next room.

Remote tests were carried out in participants' work settings. Participants' telephone set was replaced by the test telephone during the test sessions. Although replacing the telephone set was an easy task for the experimenter, setting up the observation hardware and software without invading the participants' personal environment was the biggest challenge for the experimenter. So as to minimally invade participants' physical space it was decided to utilize a mini USB web-cam with 800x600 image resolution to record the usage data. The web-cam was connected to users' own computer. Monitoring software was able to detect motion and then trigger the recording for a given time.

Contrary to the simulated tasks of laboratory tests, remote participants carried out their daily tasks during the tests. Consequently, recording sound during remote tests raised some privacy issues as a result sound was not recorded during remote tests.

Protocol: A debriefing session was held before each test session both in remote and local settings. A pre-test and post-test questionnaire was given to all participants to gather data on demographic information, cellular phone usage, telephone set usage and preferences on usability test setting. In remote settings due to privacy issues, camera was adjusted to capture only the telephone set; as a result experimenter was unable to see participants' facial expressions during usage.

Table 1, Test protocols for each setting

Laboratory Testing	Remote (Field) Testing
	Equipment setup
Debriefing	Debriefing
Pre-test questionnaire	Pre-test questionnaire
Test session (45 min. in average)	Test Session (6 hours in average)
Think aloud protocol with audio/video recording	Critical incident reporting and task logging on paper, video recording with motion sensitive web-cam.
Post-test questionnaire	Post-Test questionnaire

During the study all users performed the same five tasks. In addition to the given tasks remote users were free to use the telephone set for their daily activities. In remote tests the experimenter visited the participants at 9:30 in the morning, set-up the hardware and software and then conducted the debriefing in half an hour. Remote test sessions were started at 10:00 am and finished at 16:00 pm. In contrast to the remote tests laboratory test sessions took 45 minutes per subject on average. Local test sessions were completed in two days whereas remote tests took 5 days in total. A brief comparison of both test protocols, remote and laboratory, is listed in Table 1.

Data Gathering and Analysis: In the laboratory tests all participants were given five task scenarios on frequently used and basic functions of the telephone set. The tasks were; *call redirect, make a call from address book, save an entry to address book, check missed calls and assign a call diverting shortcut to a function button.* Task difficulty gradually increased from Task 1 to Task 5, Task 1 being the simplest one. Although remote users were free to use the telephone at their will, the same 5 tasks were also given to them to form a basis for comparing the two evaluation methods.

During remote test sessions, inspired by the 'critical incident method' (Hartson & Castillo, 1998) in software interface usability evaluation, participants were asked to report the problems they faced during the tasks on separate papers with time stamp. Hartson and Castillo (1998) defined critical incident as *'an occurrence during user task performance that indicates something (positive or negative) about usability'.* As sound was not recorded in remote settings, these incident reports were quite useful for the evaluator in identifying the exact time of the incident and additional attention was paid to those moments in the video files. Moreover, to some extent, incident reports replaced the 'think-aloud protocol' which is broadly utilized in laboratory -based usability activities.

Findings and Discussion

As gaining insight and experience in remote testing rather than searching for statistical significance was the major goal, task completion rates (Table 2), number of usability problems (Table 4) and preference of subjects on test method (Table 5) were gathered to create a common ground for comparison.

As the difficulty of the tasks was increased gradually from easier to harder, the first task being the easiest, a task based comparison was utilized. An "independent samples t test" was conducted to evaluate the hypothesis that there is no significant difference between remote and local subjects in terms of task completion rates (Table 2) during the usability test.

Table2, Task completion rates of the users. 1 stands for a completed, 0 for not completed task and 0,5 is for a semi completed (some certain criteria satisfied) task.

	user	Task 1	Task 2	Task 3	Task 4	Task 5
Remote	R1	1	0,5	1	0	0
	R2	1	1	1	0	1
	R3	1	0,5	1	0	0,5
	R4	1	0	0	0	0
	R5	1	0	0	0	0
Local	L1	1	0	0	0	0
	L2	1	0,5	0	0	0
	L3	1	0,5	1	0,5	0
	L4	1	0,5	0,5	0	0
	L5	0	0,5	1	0,5	0

The equality of means in task completion rates might be because of the contextual factors, 'distractions' in remote settings. Although, remote subjects tested the product for extended periods of times, local subjects used the product in a more isolated, concentrated and committed manner. The results of the t test are listed in Table 3.

Table3, Results of independent samples t test of local and remote groups in terms of task completion rates (α=0,05).

		Mean	Std Dev.	t	P
Task 1	Local	0,8	0,4472	-1	0,347
	Remote	1,0	0,0000		
Task 2	Local	0,4	0,2236	0	1
	Remote	0,4	0,4183		
Task 3	Local	0,3	0,4472	-0,949	0,371
	Remote	0,6	0,5477		
Task 4	Local	0,2	0,2739	1,633	0,141
	Remote	0,0	0,0000		
Task 5	Local	0,0	0,0000	-1,5	0,172
	Remote	0,3	0,4472		

5 predefined tasks which were expected to occur frequently in real world usage were given to both remote and local test participants. In remote test setting, test subjects were required to complete the same tasks but they were free to use the telephone at their convenience in their daily routine. After the evaluation of test sessions, 22 usability problems were identified in total, of which 14 were common in both settings, 3 were observed in local and 5 were observed in remote settings (Table 4). In remote test setting due to the contextual factors in use, test subjects tried other functions of the telephone. These additional data resulted in the identification of 7 new usability problems. In line with the literature (Thompson, Rozanski & Haake, 2004, Tullis et al. 2002) remote tests identified more usability problems in total. Moreover, remote subjects provided some key usability findings regarding the context of use. At one of the remote test sessions at a local radio station, the telephone set was in the broadcasting studio and the participant tried to set the telephone to silent mode. During these trials evaluators observed 3 usability problems. Another contextual factor that surprised the researchers was that the need for using the telephone under strict time limit put the participant under big pressure.

Table 4, Usability problems identified during the tests

	Remote/Field	Common	Laboratory	Total
Task based	5	14	3	22
Context based	7	--	--	7
Total	12	14	3	29

A post test questionnaire was distributed to all subjects to gather information on their experience on the test method. Subjects were asked to mark their preference in a 5 degree scale from strongly disagree (1) to strongly agree (5). The responses to related items (Table 5, Q5, Q6 and Q7) in the questionnaire showed that participants at both settings preferred and would prefer to carry out usability testing at their workplace rather than a usability laboratory. This was not an unexpected result for the evaluators as it is in line with software testing literature (Brush, Ames & Davis, 2004). The reasons behind this preference might be as follows:

- The convenience of taking the test without leaving the workplace.
- The ability to continue daily work during the test.
- Avoiding the pressure of being observed by cameras and the one-way mirror in the usability laboratory.

However, the responses given to questions related with observational space (Table 4, Q1, Q8, Q10, Q12, Q3 and Q14) show that test subjects seemed comfortable while being observed and videotaped.

Table 5, Responses given to the post test questionnaires on 5 point Likert scales running from 1= strongly disagree, 5 = strongly agree].

			Remote (mean)	lab (mean)
Common	Q1	Being observed during the usability test did not make me anxious	4,2	4,2
	Q2	I prefer to be observed by a person instead of a camera	3,8	2,8
	Q3	It would have been better if there was a person with expert knowledge on the telephone, with me during the test sessions.	4,2	3,2
	Q4	I feel lost during the test session	2	3
	Q5	I prefer to carry the test at my workplace.	4,2	3,2
Remote	Q6	It would be an annoying experience to conduct the test at an usability laboratory	1,2	
	Q7	I prefer attending to test at the usability laboratory to attending from my workplace	2,6	
	Q8	I would be annoyed if the entire environment was recorded instead of the product interface	2,6	
	Q9	I feel better by seeing and checking the recorded image from my computer	3,2	

	Q10	I will be annoyed if there are more cameras and I do not have the chance to check the recorded image.	2,6	
Local	Q11	It was an annoying experience to conduct the test at usability laboratory		1,8
	Q12	I would have feel more comfortable if just the product interface was recorded instead of the entire environment.		2,2
	Q13	I would have felt more comfortable if I had known what has been recorded.		2
	Q14	I feel annoyed by being observed by a couple of cameras.		1,8

Conclusion

The present study tried to compare laboratory and remote product usability testing. Even though this was an initial inquiry of an exploratory study, it was found that usability testing and evaluation of a product, with users' own tasks and goals, in actual use context, reveals implicit usability problems in the interface. Moreover, extended periods of test sessions relative to laboratory based testing, real tasks and goals instead of simulated ones provided valuable usability information on products' usability on real use contexts and on users' learning and retention of product interface.

Although users reported that they feel comfortable while being video and audio taped during a laboratory based usability test, the post test questionnaires showed that they would prefer to participate to the usability tests from their workplaces.

References

Ames, M. and Brush, A.J. (2003). Final report on remote usability studies. CRA-W Distributed Mentor Project. Retrieved September 9, 2005 from www.cra.org/Activities/craw/dmp/awards/2003/Ames/report.html)

Brush A.J.B,Ames M. and Davis J. (2004). A Comparison of Synchronous Remote and Local Usability Studies for an Expert Interface. *Proceedings of CHI 2004*, 1179-1182

Dray S. and Siegel D. (2004). Remote Possibilities? International Usability Testing at a Distance. *Interactions*. Vol 11 (2), 2004, 10-17

Gough, D. and Phillips, H. (2003). Remote Online Usability Testing: Why, How, and When to Use it." Retrieved September 9, 2003 from http://www.boxesandarrows.com/archives/remote_online_usability_testing_why_how_and_when_to_use_it.php

Hartson, H. R, Castillo J. C, Kelso J. & Neale W. C. (1996) Remote Evaluation: The Network as an Extension of the Usability Laboratory. *Proceedings of CHI 1996*, 228-235

Hartson, H. R. and Castillo J. C. (1998). Remote evaluation for post-deployment usability improvement. *Proceedings of the working conference on Advanced Visual Interfaces* May 1998 L'Aquila, Italy. pp 22-29

Kantner L, Sova H. D and Rosenbaum S. (2003) Alternative Methods for Field Usability Research. *Proceedings of SIGDOC 2003*, 68-72

Petersen M.G,Madsen K.H, and Kjaer A. (2002). The Usability of Everyday Technology—Emerging and Fading Opportunities. *ACM Transactions on Computer-Human Interaction*, Vol. 9, No. 2, June 2002, 74–105.

Rowley, D. E. (1994). Usability testing in the field: bringing the laboratory to the user. *Proceedings of the SIGCHI conference on Human factors in computing systems*: *celebrating interdependence*, 252-257.

Thomas P. and Macredie R. D. (2002). Introduction to New Usability. *ACM Transactions on Computer-Human Interaction*, Vol. 9, No. 2, June 2002, 69–73.

Thompson K.E, Rozanski P.E & Haake A.R. (2004). Here, There, Anywhere: Remote Usability Testing That Works. *Proceedings of SIGITE'04*, October 28-30, 2004, Salt Lake City, Utah, USA. 132-137

Tullis, T., Fleischman, S., McNulty, M., Cianchette, C., and Bergel, M.(2002). An Empirical Comparison of Lab and Remote Usability Testing of Web Sites. *Proceedings of Usability Professionals Association Conference*, July 2002. Retrieved June,2005 from http://home.comcast.net/~tomtullis/publications/RemoteVsLab.pdf

Wichansky A.M. (2000). Usability Testing in 2000 and Beyond. *Ergonomics*, 2000, Vol. 43, No. 7, 998-1006