

Rochester Institute of Technology

RIT Digital Institutional Repository

Theses

2012

What is good enough for mobile system evaluation: A Comparison of laboratory and field setting

Yao Wang

Follow this and additional works at: <https://repository.rit.edu/theses>

Recommended Citation

Wang, Yao, "What is good enough for mobile system evaluation: A Comparison of laboratory and field setting" (2012). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by the RIT Libraries. For more information, please contact repository@rit.edu.

What is Good Enough for Mobile System Evaluation: A Comparison of Laboratory and Field Setting

By

Yao Wang

Thesis submitted in partial fulfillment of the requirements for the
degree of Master of Science in Human Computer Interaction

Rochester Institute of Technology

**B. Thomas Golisano College
of
Computing and Information Sciences**

5/24/2012

Rochester Institute of Technology

**B. Thomas Golisano College
of
Computing and Information Sciences**

Master of Science in Human Computer Interaction

Thesis Approval Form

Student Name: Yao Wang

Thesis Title: What is Good Enough for Mobile System
Evaluation: A Comparison of Laboratory and Field Setting

Thesis Committee

Name	Signature	Date
<u>Evelyn Rozanski</u> Chair		<u>5/24/2012</u>
<u>Michael Yacci</u> Committee Member		<u>5/24/2012</u>
<u>Anthony Jefferson</u> Committee Member		<u>5/24/2012</u>

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	5
ABSTRACT	6
LIST OF TABLES	7
LIST OF FIGURES	8
1 INTRODUCTION	9
1.1 <i>Field Evaluation Is Necessary, But Insufficient And Difficult</i>	9
1.2 <i>The Lack Of Consensus On Evaluation Techniques For Mobile Phones</i>	11
1.3 <i>Proposed New Study On Contrasting Laboratory And Field Evaluations</i>	13
2 RELATED WORK.....	14
2.1 <i>Simulating Real-Context In The Laboratory Environment</i>	14
2.2 <i>Data Collection Methods On Mobile Devices</i>	15
3 METHODS	17
3.1 <i>Problem Statement</i>	18
3.2 <i>System</i>	18
3.3 <i>Laboratory Environment</i>	19
3.3.1 <i>Setting</i>	19
3.3.2 <i>Participants</i>	20
3.3.3 <i>Tasks</i>	21
3.3.4 <i>Test Procedure</i>	22
3.4 <i>Field Environment</i>	23
3.4.1 <i>Setting</i>	23
3.4.2 <i>Participants</i>	24

3.4.3	<i>Tasks</i>	25
3.4.4	<i>Test Procedure</i>	26
3.5	<i>Data Collection</i>	26
4	DATA ANALYSIS & FINDINGS	28
4.1	<i>Task Effectiveness</i>	28
4.2	<i>Efficiency</i>	30
4.3	<i>Ease To Use</i>	32
4.4	<i>Usability Problems Identified</i>	33
5	DISCUSSION	37
5.1	<i>Differences Between Two Testing Environments</i>	37
5.2	<i>Simulations In The Laboratory</i>	40
5.3	<i>Users' Backgrounds Affect The Results</i>	42
5.4	<i>Recording Tool (Shootme Application) Evaluation</i>	44
6	LIMITATION	45
7	CONCLUSION	45
7.1	<i>Contributions</i>	45
7.2	<i>Lessons learned</i>	47
8	REFERENCES	48
	Appendix A1 Recruitment flyer	53
	Appendix A2 Screener	54
	Appendix A3 Informed consent	56
	Appendix A4 Orientation script	59
	Appendix A5 Usability problems identified in the laboratory	73
	Appendix A6 Usability problems identified in the field	75

ACKNOWLEDGEMENTS

Thank you to my thesis committee – Evelyn Rozanski, Michael Yacci, and Anthony Jefferson. Your continuous support and guidance proved itself paramount to my thesis process. I am forever grateful to have worked with and learned from you.

I want to thank Brian O’Keefe, for providing helpful comments to previous version of my thesis.

I also want to thank all of the student participants for this study.

I also wish to extend my gratitude to my parents and my friends– I am sincerely indebted for your encouragement, advice, assistance, and friendship.

ABSTRACT

There is no consensus on whether field tests are necessary in evaluating mobile systems. Empirical studies have compared evaluations in the laboratory environment and field environment, but they provide different results due to the lack of consistent testing environments. Contributing to mobile system evaluation exploration, this study conducts a comparative analysis between laboratory study and field study by recreating the natural context in the laboratory and using recording software embedded in the mobile device. The results show that involving natural use context in a controlled testing environment identifies more usability problems than a complex field test, when the objective is to uncover mobile system design details. If the goal of a usability evaluation is to involve users' natural behaviors and examine context-dependent usability problems, a field test is necessary. We further analyze how to effectively simulate the natural context in the laboratory environment. The conclusion shows that the results might be different when using different user pools.

LIST OF TABLES

Table 1. Laboratory user profile.....	21
Table 2. Contextual scenarios in the laboratory.....	22
Table 3. Field user profile.....	25
Table 4. Users' phone types.....	25
Table 5. Contextual scenarios in the field.....	26
Table 6. Subjective measures on ease to use	32
Table 9. Severity of usability problems identified.....	36
Table 10. Critical usability problems.....	37

LIST OF FIGURES

Figure 1. Android mobile system interface.....	19
Figure 2. Lab testing environment	20
Figure 3. Field testing environment.....	24
Figure 4. ShootMe application interface	27
Figure 5. Success rate in the two testing environments	29
Figure 6. Mean time spent on tasks successfully completed.....	31
Figure 7. Unique usability problems identified by users in the laboratory.....	34
Figure 8. Unique usability problems accumulated by users in the field	35
Figure 9. Users are affected by the context factors	41

1 INTRODUCTION

1.1 Field Evaluation Is Necessary, But Insufficient And Difficult

Nowadays, the study of usability evaluations for mobile devices is becoming an emerging research area in Human Computer Interaction. Many studies have been done to help mobile devices adapt to changes in motion by increasing the font size or other device parameters, but rules for determining when and how to design the device are lacking due to the complex of field testing under changing environment (Barnard, Yi, Jacko & Sears, 2005). Pascoe, Ryan and Morse (2000) pointed out the portable computing device is very different because of its inherent mobility; it allows users to work with their data whilst on the move and considers it as context-dependent. Some other studies also found that performing tasks while walking requires more attention and can lead to slow task performance, as compared to standing and sitting (Barnard et al., 2005). Johnson(1998) has been concerned that laboratory evaluations do not simulate the real context and lack the desired ecological validity. As a result, it is assumed that more attention should be paid to situational considerations in evaluating mobile devices such as interaction style, privacy, mobility, context, lighting conditions, varying levels of noises, multitasking as well as unexpected external interruptions (Kjeldskov, Skov, Als & Hoegh, 2004, Sá & Carriço's, 2008, Lee & Grice, 2004, Väättäjä & Roto, 2010).

However, only 41 percent of mobile usability papers are empirical in nature (Kjeldskov & Graham, 2003), it is much more difficult to use evaluation techniques and data collection methods in the natural settings (Kjeldskov et al., 2004). Control

is critical to collect empirical data and make the data interpretable and accurate, but a field setting is an uncontrolled testing environment (Lee et al., 2004, Barnard et al., 2005). Moreover, field-testing is considered as a time-consuming exhaustive analysis compared with other approaches (Kjeldskov et al., 2004, Duh, Tan & Chen, 2006, Kjeldskov, Graham, Pedell, Vetere, Howard, Balbo & Davies, 2005), and Kaikkonen, Kallio, Kekalainen, Kankainen and Cankar (2005) even found that testing in the field requires double time in comparison to the laboratory. Another big obstacle for conducting usability tests on mobile devices in the real use context is the lack of supporting tools. Schusteritsch, Wei and LaRosa (2007) mentioned that enabling natural interaction with the device is more challenging because of the small screen size, form factors and diversity of phone shapes. Video equipment to be carried by a follower is the most popular method used in the field to enable the natural interaction (Kjeldskov et al., 2005, Kjeldskov & Stage, 2004), continuously shifting focus between the mobile device, the user and the surrounding environment (Kjeldskov et al., 2004). But it is impossible to collect accurate data since the user is moving all the time and users' performance data on interacting with mobile phone might be lost. To improve this method, attaching an external wireless camera that can be mounted on the mobile phone is designed by some usability practitioners (Nielsen, Overgaard, Pedersen, Stage & Stenild, 2006, Duh et al., 2006, Kaikkonen et al., 2005, Sá & Carriço, 2008, Betiol & Cybis, 2005). But this method is quickly found as problematic because of the heavy weight of the entire unit, unsecure mount, and easily distracting users' attention (Schusteritsch et al., 2007). There is also another concern that the added value in the field is unknown, so trying to avoid a complex

time-consuming field study on mobile devices seems plausible, but in fact it is questionable.

1.2 The Lack Of Consensus On Evaluation Techniques For Mobile Phones

As a result, HCI practitioners are provided with little support on gaining consensus on evaluation techniques for mobile phones. Lee and Grice (2004) proposed a combination method for testing mobile application by applying the scenario-tasks, questionnaires and heuristic evaluation techniques, which introduced the new rules for how to combine testing methods in order to design and develop mobile applications. More studies have been conducted by comparing different usability testing approaches to establish formal appropriate mobile device evaluation methods. Kjeldskov et al. (2005) compared the four approaches in a mobile device evaluation study including field evaluation, laboratory evaluation, heuristic walkthrough and rapid reflection. By comparing the results of laboratory and field testing, Kjeldskov et al. (2005) pointed out that in the laboratory environment many problems were related to interface issues while the field study stressed problems of mobile use rather than simply device usability and concluded that at least we should collect broader data in the field. Duh et al. (2006) also pointed out the laboratory usability test was insufficient in evaluating the mobile device in its actual context of use. By contrasting the frequency and severity of usability problems users encountered in laboratory environment and field setting, significant differences were uncovered in this study; and several of the critical problems were only associated with the device being used in the field. This conclusion was also

confirmed by Nielsen et al. (2006), who found testing mobile devices in the field valuable. Although usability issues related to the interface were similar in these two environments, in the field-based evaluation they did identify significantly more usability problems involving interaction style and cognitive load that were not identified in the laboratory.

However, some other similar comparative studies, of which results were to the contrary, lead to an intense argument about whether conducting field tests on mobile devices are worthwhile. Several comparative studies (Kaikkonen et al., 2005, Kjeldskov et al., 2004) conducted both in field and laboratory environments indicated that there is little added value of conducting a complex field test when for searching user interface flaws. Kjeldskov and Stage (2004) also agreed with this conclusion indicating that seating the test subjects at a table identified significantly more usability problems than any of the other proposed technique. Kaikkonen et al. (2005) found that potential interruptions in the field did not seem to affect the performance much and concluded that laboratory tests give sufficient information when testing usability of a mobile application. But they also mentioned that when user behavior is investigated in a natural context, the field study might be worthwhile. Betiol and Cybis(2005) also indicated that many important usability problems can be found in simpler laboratory approaches. Their study also suggested this experiment should be repeated using real phones in these two different contexts with the user standing or walking. Similarly, Kjeldskov's (2004) results showed that recreating central aspects of the use context in a laboratory setting revealed more usability problems than the field setting.

1.3 Proposed New Study On Contrasting Laboratory And Field Evaluations

By comparing the laboratory experiments and field studies from a research perspective, the former method seems an environment with controlled variables like mentioned before but lack of realism, while the latter method contributes more in the natural user behavior but is difficult to collect data (Kjeldskov & Graham, 2003, Johnson, 1998). Most laboratory settings were traditionally quiet and comfortable laboratory environments for users to concentrate on the tasks, the usability problems that were not identified in the laboratory were assumed as highly context-dependent. Also, when examining the users' subjective behaviors and field environments, field-testing may add significant value to the results. Apparently, it has been suggested that instead of going into the field, adding contextual features to laboratory settings can be the tradeoff between maintaining the benefits of a controlled environment and involving the context factors that might affect users' performance (Kjeldskov et al., 2004, Barnard et al., 2005). However, only 114 of 636 papers deal with evaluating a mobile system, even worse, only 6 out of the 114 papers laboratory evaluations involve new techniques being applied to meet the challenges. (Kjeldskov & Stage, 2004, Duh et al., 2006)

In the light of this, our study aims to strike the balance between laboratory studies and field studies by applying new techniques to contrast the results. To address some of the shortcomings of past studies, our study: (1) introduces a new application to capture the mobile screen directly that is installed internally on a mobile device and provides high-quality video data for further analysis, and (2)

applies more key factors like noises, distractions and unexpected interruptions that might influence users' performances to improve the realism of a laboratory setting. Very few studies have done this before, only Kjeldskov et al. (2004) and Kjeldskov & Stage (2004) have tried to simulate the real context in the laboratory environment, but both of their studies only targeted on the mobility and context. To add more value in our study, we will include more scenario variables such as noises, distractions, unexpected interruptions and changing light condition in the laboratory testing environment.

2 RELATED WORK

2.1 Simulating Real-Context In The Laboratory Environment

There were mainly two methods in adding real context into the laboratory setting: one method used a treadmill, pre-defined track or real world route to simulate walking conditions and the other method used a car simulator to simulate driving scenarios (Duh et al., 2006). Kjeldskov and Stage (2004) also tried to explore new techniques to simulate mobility in laboratory settings, which included two frameworks: one was related to body motion like walking on the treadmill or walking on a track with obstructions changing all the time; the other one was based on the notion of divided attention, which is divided between physical motion navigation and the use of the system. Similarly, in Kjeldskov et al.'s (2004) study, the usability laboratory was set up to resemble a part of the physical space of a hospital department, which extended the standard experiment setup to include mobility and context. There were two evaluation rooms, connected by a hallway, which were

furnished with beds and tables similar to real hospital wards. The test subjects were required to walk between the two rooms. Both these two studies only focused on mobility and environment. However, Barnard et al. (2005) conducted another interesting study, which contributed additional rigor by employing two light levels in testing scenarios, that suggested a controlled walking scenario would adequately simulate the actual user experience. This study designed two testing environments in the same room; one used a treadmill to simulate a normal walking scenario at a fixed speed and another used a controlled walking scenario, furniture in the room were considered as obstacles, and users were allowed to walk at varying speeds following a predefined path, but they had to keep moving. The drawback of this experiment is that all these simulations are only discussed within the laboratory environment and lacked the comparison with the real context.

It is generally recognized that other factors like noises, distractions and interruptions in the field also might affect users' performance. A lot of studies pointed out these important issues (Jensen, 2007, Sá &Carriço, 2008, Duh, 2006), but none of the studies tried to simulate these important factors in the laboratory environment. Kjeldskov and Graham (2003) also suggested that evaluations often focused on interface functionality rather than contextual issues. So in this comparative study, noises, different levels of lighting, unexpected interruptions, as well as mobility; contexts are all applied to simulate the laboratory environment to increase the realism.

2.2 Data Collection Methods On Mobile Devices

Magnusson et al. (2010) summarized several methods to gather mobile user experience data from different aspects of the mobile usage situation like interviews, focus groups, and user workshops. To collect preference data, our study will focus on the answers given on the move and questionnaires after each task.

When compared to collecting user preference data, capturing user performance data is challenging in the field due to the high mobility of mobile systems and the lack of supporting tools (Schusteritsch et al.,2007). Video equipment carried by an observer is the most popular method used in the field (Kjeldskov &Stage, 2004, Kjeldskov et al., 2005), continuously shifting focus between the mobile device, the user and the surrounding environment. But it is impossible to collect accurate data since the user is moving all the time and it is easy to lose the users' performance data on interacting with the mobile interface. To collect more complete data sets, an external wireless camera is mounted on the mobile phones (Betiol & Cybis, 2005, Duh et al., 2006, Kaikkonen et al., 2005, Nielsen et al., 2006, Sá & Carriço, 2008). This method is quickly determined as problematic because of the heavy weight of the entire unit, the unsecure mount, and the distractions to the users' attention (Schusteritsch et al., 2007). To capture the screen naturally and directly, some studies (Hong et al., 2001, Waterson et al., 2002) developed a technical solution involving software embedded in the system. The most popular tool is called WebQuilt (Hong et al.,2001),which is designed for remote usability testing. Logging is done through a proxy. Waterson et al. (2002) did a comparative study between a traditional laboratory set-up using video cameras and a field-based test using WebQuilt to collect usage information. This method of using a proxy-based logging

and analysis tool to gather clickstream data involves a time-consuming preparation process and difficult data interpretation, and only captures the right click events, so actions such as clicks on non-clickable areas or clicks on the screen before the page was fully loaded were not able to be recorded, other data like scrolling and text input were totally lost. Liisa and Katja(2010) have reported their findings from user tests performed with an eye-tracking system in a forest environment. Using eye-tracking systems in a mobile context is valuable in improving the methods of collecting mobile user experience data. The most challenging issue right now in using an eye-tracking system is the poor data quality (Liisa&Katja, 2010); three-dimensional environments and the vagaries of lighting conditions outside will influence the calibration of the eye-tracker, so gaze data may not be trustworthy enough especially on the mobile device's small screen.

As a result, the only feasible approach is to use screen capture software, which is similar to the ones used for the desktop. Balagtas-Fernandez and Hussmann (2009) considered finding a screen capture tool as a challenge due to the privacy issues and security limitations posed by the mobile device. In our study, we propose a new screen capture tool, which is similar to a log system, but can record the main interface and save the video clips to the phone memory card.

3 METHODS

Two user-based usability evaluations were conducted. All of the test materials were identical, only the test environments were different. The laboratory group used a

conference room, while the field group evaluation was conducted in the field on a college campus. Both of the tests followed Rubin's (2008) usability testing guidelines.

3.1 Problem Statement

Our comparative study aimed to explore the similarities and differences between usability tests conducted in the laboratory and in the field with regard to quantitative measures such as number of usability problems identified, severities of usability problems, task success rates and task execution times. We also focused on qualitative descriptions such as users' behaviors, users' interaction styles and users' subjective comments.

3.2 System

The Android operating mobile system was selected (Nauman et al., 2010). Since Motorola had a significant share of the smartphone market ("PCworld", n.d.), we chose the Motorola Bravo with Android platform 2.2 (Figure1). The desktop functions were customized for making calls, editing contacts, showing applications, checking emails and searching online. Feedback such as WIFI signal connection and battery life were shown on the top toolbar. A list of applications, which was required when performing the tasks, was pre-installed on this mobile device. To enable the screen capture tool to start the recording server, this mobile phone was rooted, a special process that provides root privileges to the user and leads to attaining control with Android's Linux subsystem.

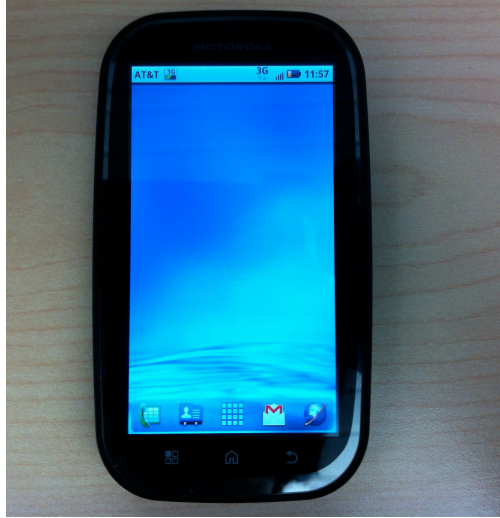


Figure 1. Android mobile system interface

3.3 Laboratory Environment

3.3.1 Setting

The laboratory evaluation was conducted in a typical controlled environment; the conference room was used instead of the traditional laboratory in this study so that the space was big enough to simulate the contexts (Figure2). Tables and chairs were considered as obstacles, and a laptop was positioned on the table to make some noise. The participant was physically mobile in the room when doing some tasks, and the moderator followed the participant to make a direct observation.



Figure 2. Lab testing environment

3.3.2 Participants

Previous research showed that four or five participants would reveal approximately 80% of the usability problems in any product (Duh et al., 2006). Rubin and Chisnell (2008) suggested that at least eight participants of each type be tested to expose as many usability problems as possible in the shortest amount of time.

Eight college students (three males and five females) aged from 18-34 years old were recruited. All of eight participants were currently using mobile phones and the phone types included iPhones, Androids, BlackBerrys and non-smartphones. All of them indicated that they always remember to bring their mobile phones every day and described themselves as experienced phone users. The eight users were equally distributed with four different phone types. Six of the eight indicated that making phone calls and sending messages were the most frequently used functions. Four of the eight considered checking email as their most frequently used function. Three of

the eight responded that they usually used their phones to search for specified information. A detailed users profile is shown below, in Table 1.

Demographic Characteristics	Participant number (N=8)
Gender <ul style="list-style-type: none"> • Male • Female 	 3 5
Age <ul style="list-style-type: none"> • 18-21 • 22-24 • 25-34 	 4 3 1
Phone type <ul style="list-style-type: none"> • Android • IPhone • Blackberry • Non-smartphone 	 2 2 2 2
Phone use frequency <ul style="list-style-type: none"> • Always remember to bring it everywhere everyday 	 8

Table 1. Laboratory user profile

3.3.3 Tasks

Task scenarios for mobile system evaluations were composed of a set of variables related to social environment, movement and posture, device characteristics, mental workloads and user personas (Sá & Carriço's, 2008).

In our study, contextual scenarios in the laboratory were defined involving noises, obstacles, movement transitions, levels of light, and interruptions. The tasks were associated with the primary functions of Android system version 2.2, focusing on daily usage functions like making phone calls and sending messages, as well as smartphone primary functions, such as checking email, dealing with applications,

and searching online for specified information. The detailed tasks are shown in Table 2.

Tasks	Contextual variables
Task1. Download a specified app	Keep walking around the obstacles
Task2. Send a specified SMS	Move from one chair to another chair
Task3. Edit a phone number	Sit in the conference room with lots of noise
Task4. Reply to a specified email	Be interrupted and required to search a song during the task
Task5. Uninstall a specified app	Respond to an unexpected call during the task
Task6. Search for nearby restaurants on maps	Walk in an environment with lights switching between low and high luminance

Table 2. Contextual scenarios in the laboratory

3.3.4 Test Procedure

Six tasks, with a predefined order, were given to each of the eight participants. The first two tasks were always provided in a set sequence because of the walking requirement while the last four tasks were counterbalanced. The procedure consisted of four parts:

1. A brief orientation and a pre-test questionnaire; the orientation introduced the study and the questionnaire gathered users' background information.
2. An exercise task; the participant was asked to learn the mobile phone interface while they were sitting at a table.

3. Tasks were given to the participants. At the end of each task, a quick interview was held with participants to clarify their interactions and a post-task questionnaire was given.

4. A post-test questionnaire and a debriefing; was used to obtain overall experience from participants.

3.4 Field Environment

3.4.1 Setting

The field evaluation, which was focusing on the system use in natural settings, was conducted on the Rochester Institute of Technology campus. The participants were asked to do whatever they normally would do while performing the phone tasks.

There were always other people in the testing area. The participant was instructed to walk following a predefined routine and the moderator followed the user a few steps behind to give the tasks and make direct observations. Parts of the tasks were finished when they were moving. Parts of the tasks were accomplished in a natural context. The field environment was a predefined path on campus including a pedestrian street in a parking lot, a set of stairs and a noisy dining hall (Figure 3).



Figure 3. Field testing environment

3.4.2 Participants

A second set of eight participants (six males and two females) aged from 18-34 was recruited on campus. The participants were similarly distributed as iPhone users, Android users, BlackBerry users and non-smartphone users. All of the participants indicated that they always remember to bring their mobile phones every day and described themselves as experienced phone users. The participants in this group were also equally distributed with phone types. . Four of the eight indicated making phone calls, sending messages and searching for specified information were the most frequently used functions. Three of the eight answered that they also frequently used their mobile phones to check email. A detailed users profile is shown below, in Table 3.

Demographic Characteristics	Participant number (N=8)
Gender <ul style="list-style-type: none"> • Male • Female 	6 2
Age <ul style="list-style-type: none"> • 18-21 • 22-24 • 25-34 	4 1 3
Phone type <ul style="list-style-type: none"> • Android • iPhone • Blackberry • Non-smartphone 	2 2 2 2
Phone use frequency <ul style="list-style-type: none"> • Always remember to bring it everywhere everyday 	8

Table 3. Field user profile

The participants' backgrounds indicating the phone types of the both the field group and laboratory group are listed in sequence below (Table 4).

Participant#	Lab users	Field users
1	Blackberry	IPhone
2	Non-smartphone	Non-smartphone
3	Android	Android
4	Non-smartphone	Android
5	IPhone	Non-smartphone
6	Android	Blackberry
7	Blackberry	IPhone
8	IPhone	Blackberry

Table 4. Users' phone types

3.4.3 Tasks

Tasks in the field were also divided into two parts: the assignments related to the mobile interface were identical; the contextual variables were natural environment

in the field, which were similar to the laboratory simulations. The detailed tasks are shown in Table 5.

Tasks	Contextual variables
Task1. Download a specified app	Keep walking around the lounge area on the 2nd floor of the building
Task2. Send a specified SMS	Walk from the 2nd floor to the 1st floor in the building
Task3. Edit a phone number	Sit in the Deli area in the building
Task4. Reply to a specified email	Be interrupted by a person and required to answer a series of questions
Task5. Uninstall a specified app	Respond to an unexpected call during the task
Task6. Search for nearby restaurants on maps	Walk in the parking lot under outside natural light

Table 5. Contextual scenarios in the field

3.4.4 Test Procedure

All of the eight participants were asked to finish the six tasks in a predefined order using the same test process as in the laboratory environment. In order to include a flickering light condition and a noisy dining background, the field test took place at noon.

3.5 Data Collection

Data collection included time spent on each task, success rates, and questions regarding ease of use. Usability problems were analyzed from the performance errors and users' comments and confusion.

Performance data and preference data were collected in the same way in both evaluation environments. During the test session, the user was thinking aloud. Comments were gathered from an audio recorder. Questionnaires and brief interviews after tasks were used to collect preference data. To record high-quality close-up views of the screen, the ShootMe application, which was downloaded from Android Market and installed on the Android system, allows you to take screenshots of the device (see Figure 4) was selected, which was available for a rooted Android mobile phone.

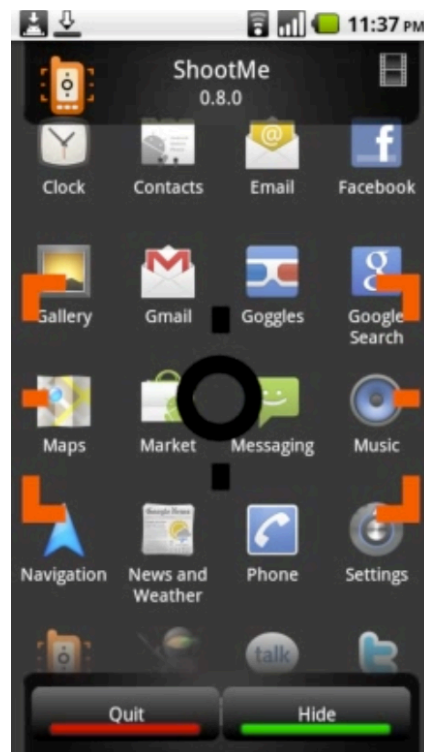


Figure 4. ShootMe application interface

Compared with several other applications, ShootMe allows a continuous screen capture feature and can record a higher resolution (up to 20 frames per second). However, no cameras were used, either in the laboratory or in the field, to maintain realism and avoid distractions.

4 DATA ANALYSIS & FINDINGS

Based on the data from background questionnaires, there were no differences between the two groups of participants. Comparative analysis between laboratory and field is described from the perspectives of effectiveness, efficiency, ease to use and usability problem identifications.

4.1 Task Effectiveness

A task was considered as completed if the end result was consistent with a predefined solution. If the test moderator observed that the user was completely frustrated, or that the user gave up trying and commented (e.g., “I don’t know what to do”, “I have no idea about this”), the test session proceeded to the next task and the current task was considered a failure.

Participants in the field experienced higher success rates for Tasks 1, 5 and 6 than users in the laboratory environment (Figure 5). It is noteworthy that Task 5, which required users to uninstall an application, had the lowest success rate in both testing environments. The uninstall process on this mobile system is complex rather than simply depressing the application icon for several seconds. Six participants

completed Task 5 in the field while only three finished this task in the laboratory. Task 1, which is divided into two parts: find the Android Market (a pre-installed application that can be accessible from Android mobile device is used to download other specified applications) and then download the Facebook application in the Android Market, experiences the second lowest success rate. Seven users successfully finished this task in the field; the user who failed this task did not know how to find the Facebook application in the Android Market. The two users in the laboratory who failed to find the Android Market were focusing on the keywords (e.g., “apps” and “installer”).

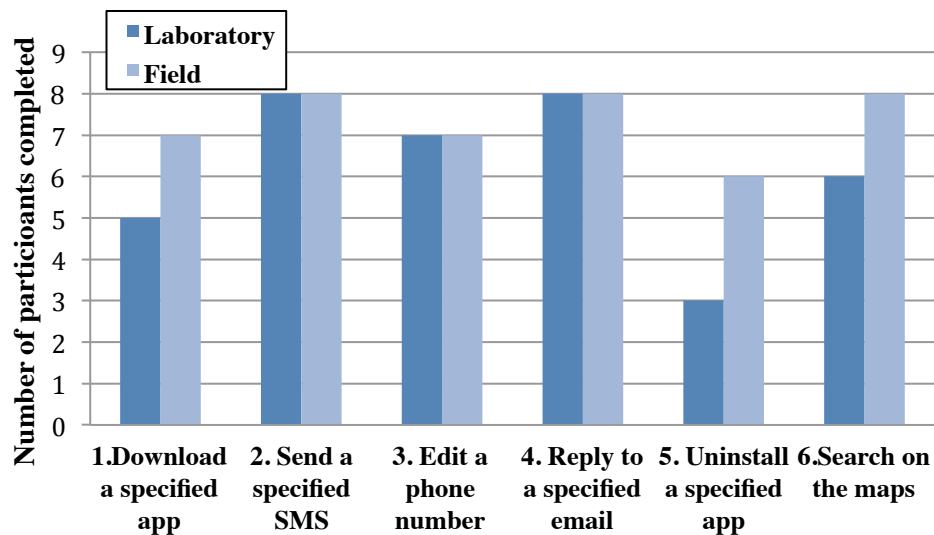


Figure 5. Success rate in the two testing environments

Task 3 was similar to the initial exercise task of becoming familiar with the system; as a result seven users completed this task in both environments. The participant that failed this task in the field experienced a longer time trying to figure out how to edit the phone number (Task 3); she wrongly called the person twice and kept

pressing the person's name expecting an edit option. It was observed that she was frustrated with several attempts, and after 238 seconds the moderator decided to offer help. The one user who failed Task 3 in the laboratory commented, "I can't find the option" and gave up trying. During the post-task interview, the user indicated that she forgot the edit button that she learned in the exercise task.

All of the participants finished Task 2 and Task 4, which focused on the main interface specific to one app (Task 2 is related to the message app and Task 4 is related to the email app). For Task 6, all of the participants in the field successfully finished this task while six of eight users in the laboratory found the desired information using the search function, and the other two users did not know what to enter into the search box.

4.2 Efficiency

Task time was measured from the time users started the task to the time users successfully completed the task. Those tasks that were not completed were not included. The execution time is calculated by subtracting downloading time, installing time, typing time and interruption time from the total performance time for each task. These values were determined by reviewing the data collected by the ShootMe tool.

For Task 1, 2, 3 and 4, users in the laboratory spent longer times than users in the field to finish these tasks. However, there is no significant difference between these two testing environments for these four tasks, even though for Task 4, users in the

laboratory took 90.37 seconds while field users took 73 seconds to successfully reply a specific email (Figure 6).

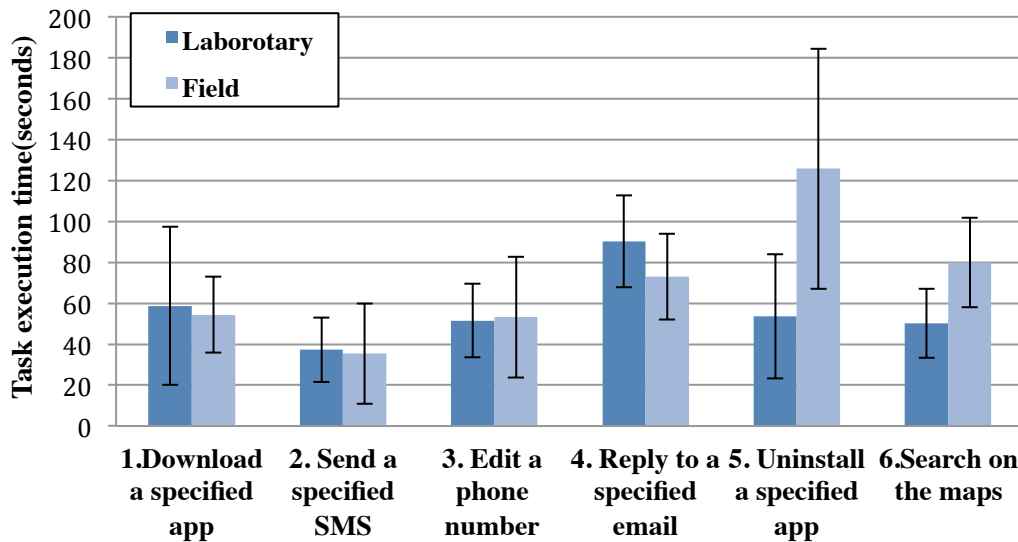


Figure 6. Mean time spent on tasks successfully completed

An independent t-test is used to compare the mean time on each task. Participants in the field experienced significant longer time on Task 5 ($t=-2.426$, $p<.05$) than users in the laboratory. The mean time spent on Task 5 in the field ($M=125.8s$) is more than double the time spent in the laboratory ($M=53.67s$), which suggests that the uninstallation on this mobile system is a complex process. Interestingly, the longer time spent on Task 5 in the field also contributes to a higher success rate in the field as shown in Figure 5.

Figure 6 also shows a significant difference on Task 6 between the two testing environments ($t=-2.889$, $p<.05$); changing light simulated in the laboratory did not significantly affect users' performance while the outside changing weather made the task much harder to finish.

4.3 Ease To Use

At the end of each task, users were asked to rate how easy it was to use this mobile device to achieve the desired goal. Each task was rated on a scale from 1 (extreme difficult) to 7 (extreme easy). Table 6 shows the detailed ratings for each task.

Tasks	Lab Mean (SD)	Field Mean (SD)	t-value	p-value
1-a. Find the desired app	3.625 (1.77)	5.5 (1.41)	-2.343	.035*
1-b. Use Search function	6 (2.13)	6.75 (0.46)	-.970	.362
2. Use message	5.875 (1.55)	5.75 (1.39)	.170	.868
3. Use phone book	5.625 (1.77)	6.25 (1.39)	-.786	.445
4. Use email	5.875 (0.99)	5.75 (0.87)	-.266	.794
5. Uninstall specified app	3.75 (2.12)	4.25 (2.19)	-.464	.650
6. Use map application	6.125 (1.13)	4.75 (1.39)	2.175	.048*

*Statistically significant at the 0.05 level.

Table 6. Subjective measures on ease to use

It is necessary to point out that users found it significantly more difficult to find the desired application (Task1-a) from the application list in the laboratory than in the field according to an independent t-test ($t=-2.343$, $p<.05$). Users were required to keep moving around the laboratory, which included several obstacles in their walking routine. While in the field, users were asked to walk around a familiar area on campus. Finding the desired application was considered as the most difficult task for users both in the laboratory (Mean=3.625) because the application list is a

long list without categories on this mobile system. It is easy to understand that users found finding a desired application along with unknown obstacles difficult.

Significant differences are also found on Task 6 (using the map application to search for nearby restaurants) ($t=2.175, p<.05$). Participants found it easier to finish the task in the laboratory than in the outside environment because of the more controlled environment. In the laboratory, the changing light did not prevent users from clearly seeing the mobile screen interface. However, in the field five of the eight participants finished this task outside with intermittent sunshine at noon, which impeded users to clearly see the mobile screen interface.

4.4 Usability Problems Identified

We identify a total of 32 different usability problems in the two testing environments. The laboratory evaluation identified 30 usability problems and the field test detected 26 usability issues. The usability issues identified in the laboratory are listed in Appendix A5 while the usability problems identified in the field are listed in Appendix A6.

Analyzing the numbers of usability issues identified in each testing environment, we found that the laboratory participants on average experience 11 problems ($SD=4.1$) and the field participants on average found 6.5 problems ($SD=2.976$). According to a Mann-Whitney U-test, participants in laboratory environment identified significantly more usability problems than users in the field setting ($U=13, z=-2.004, p<.05$).

Figure 7 shows the trend of the accumulated usability problems identified by all of the eight participants in the lab testing environments.

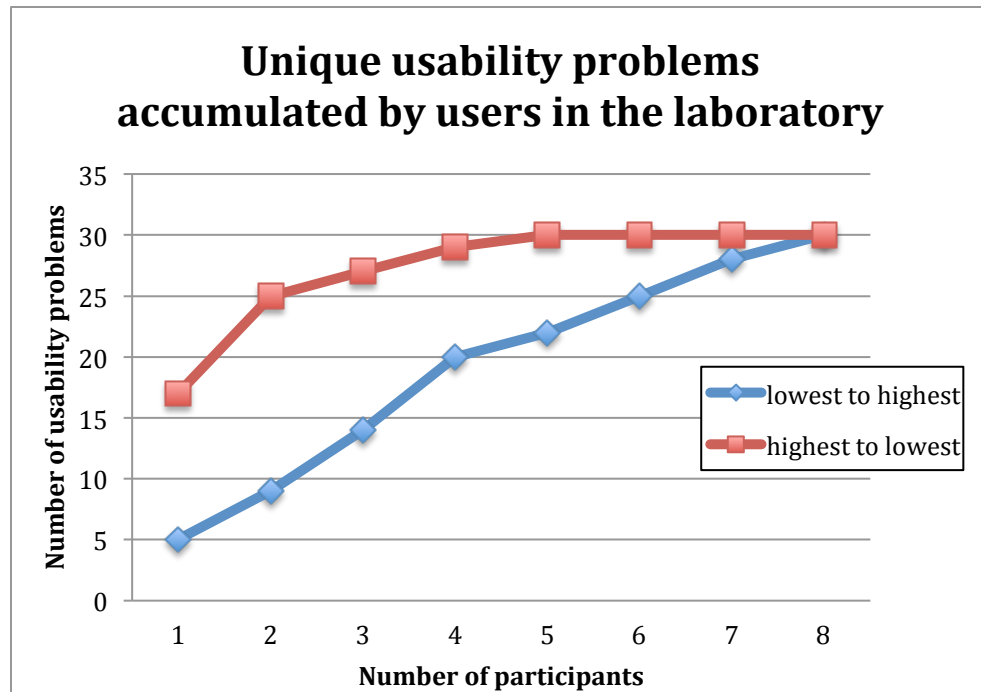


Figure 7. Unique usability problems identified by users in the laboratory

Interestingly, the trend from the highest number of usability problems that participants identified to the lowest number of usability problems that participants identified (Figure 7 red color) follows Nielsen's (1993) chart of the progress detecting usability defects in usability testing (i.e., "five is enough"). The first five participants have identified all of the 30 usability problems since the last two users are experienced users, who failed to contribute new usability problems. However, if we accumulated the usability problems from the lowest number of usability issues to the highest number of usability issues, the trend (Figure 7 blue color) is similar to linear trend. The last two participants identified additional new usability problems since they are novice users who do not use Android systems and thus were

unfamiliar with common operations. We also analyzed the unique usability problems by users identified in the field (Figure 8) using similar methods.

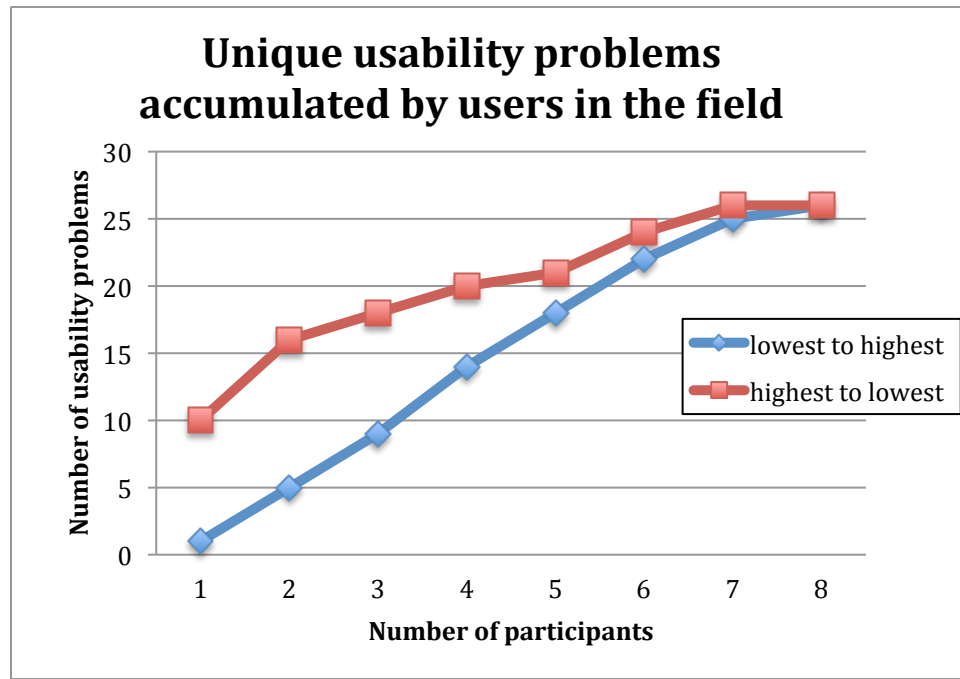


Figure 8. Unique usability problems accumulated by users in the field

Usability problems were categorized according to the criteria proposed by Molich (2000). Eight usability issues were considered as critical, 13 as severe, and 11 as cosmetic. A numeric comparison of problems between laboratory and field tests, when categorized by severity, is shown in Table 9.

The results show that the laboratory environment identified more usability problems than the field evaluation. The critical usability problems identified between the two test environments were identical. However, the laboratory environment detected four exclusive serious problems and one exclusive cosmetic problem, while the field test only uncovered two exclusive cosmetic problems. All of the exclusive usability problems identified in the laboratory were associated with

the mobile phone device, while the two exclusive usability issues identified in the field were specific to the environment; one is related to location update information, the other one is about the phone use under the sunshine.

Category	Details	Laboratory	Field	Total
Critical	Prevent users from completing tasks/ recurred across all test users	8	8	8
Serious	Increase users' time and users complete tasks eventually/recurred frequently across test users	13	8	13
Cosmetic	Increase users' time slightly but users can finish tasks easily/recurred infrequently across test users	9	10	11
Total		30	26	32

Table 7. Severity of usability problems identified

All the critical usability problems were associated with the mobile phone system, which prevented users from finishing the tasks and occurred recurrently. The detailed critical problem list is shown below, in Table 10. Except for Android users, who were familiar with the interface and system, all of the other phone type users had difficulty figuring out the process of the application uninstallation (Issue 1). Most users (14 of 16) successfully finished Task 3(Use menu button to edit users' profile), however all of them indicated that they learned to use the menu button through exercise task. Additionally, seven of 16 users reported that they did not

understand the menu button on the bottom left of the mobile phone in pre-test questionnaire before the exercise task at the very beginning of the test.

Issue#	Usability problem	Occurrence#
1	Uninstall process is hard to figure out, is not related to the application	12
2	Menu icon on the phone left bottom is not interpretable	11
3	Edit function is not related to users' profile	10
4	Hard to find the desired application from the app list	7
5	Sending email icon is hard to understand	5
6	Phone call icon is ignored	5
7	Application name is not understandable	4
8	Search icon in the Android Market is ignored	4

Table 8. Critical usability problems

5 DISCUSSION

The goal of our study was to explore the similarities and differences between testing in the laboratory and field setting. Based on the data results, the findings are presented as below:

5.1 Differences Between Two Testing Environments

We were particularly interested in the significant differences between laboratory and field by comparing usability problems, involving testing context and analyzing users' behavior pattern:

Firstly, by comparing the numbers of usability issues identified in the two testing environments, setting up a realistic laboratory uncovered more usability problems than the field. In our study, the laboratory was a controlled environment and included several factors. The noisy environment and the changing light would always occur, which contributed to finding more usability problems. While in the field, either in the Deli area or the outside weather, both the noisy context and the changing light varied. Additionally, users identified significantly more usability problems in the laboratory than in the field. Our results support the conclusion that a simple laboratory environment can identify more usability issues (Betiol& Cybis, 2005, Kaikkonen et al., 2005, Kjeldskov et al.,2004), but we failed to detect context-dependent usability problems in the laboratory. For Task 6, there is a significant difference in the average time on task between the laboratory and the field. When the user was walking far away from the building to the parking lot, the WIFI signal was weak and the smartphone switched the WIFI to 3G. Thus, the users were not sure whether the phone was still connected to the Internet since non-smartphone users were not familiar with 3G functions. One user also mentioned that since he was moving outside, the nearby restaurants should be updating. As a result, to improve the mobile system design details, a laboratory that recreates the natural context in the field can provide solid usability results. However, as Kaikkonen et al.

(2005) suggested, when evaluating applications with location information, the laboratory environment may not give the best outcome.

Secondly, users' performances are affected by a complicated context. Our study is an extension to Kjeldskov and Stage's (2004) study; we compared movement within the laboratory and field. For Task 1, users were asked to keep walking around the laboratory (room) with obstacles or around the lounge area in the field while they were searching for an application. The problem related to this particular phone system was that the application list is a long list without groupings. This complex task also revealed that users have to divide their attention between physical motion and the use of the system. We found that user familiarity with the environment determined the ease in which they finished the task. For Task 1, users found it significantly more difficult to find the Android Market application in the laboratory than in the field. All of the field users had familiarity with the outside testing area, whereas all of the laboratory users were not sure how the obstacles were positioned in the conference room. A concurrence phenomenon in both testing environments, we found that users wanted to stop searching for the specified application when they experienced difficulty and failed to complete the task anyway. Seven of 16 users mentioned that they didn't usually do complex tasks like installing when they are moving, especially with a new brand phone in a strange environment.

Finally, we found that the field was a more relaxed environment compared with laboratory. As with the findings in (Nielsen et al., 2006), laboratory participants perceived higher mental demands. During the post-task interviews, half of the field group participants reported that they were more comfortable participating in our

study than other similar usability studies, and did not consider it as a “test” because of the familiar noisy environment. One field participant who attended a traditionally usability laboratory test before mentioned that in the laboratory although he was told to do what he would normally do, he still felt nervous because of the video equipment. Users in the field experienced a significant longer time on Task 5, which led to a much higher success rate. It was observed that users in the field tried more times and experienced more errors to accomplish that difficult task. This was consistent with how they would normally accomplish this task in daily life. Moreover, although the context was simulated in the laboratory, users were concentrating more on solving the tasks; the field users were easier to be distracted by the environment. We conclude that to observe users’ behaviors and examine the environment, field testing is more suitable, which is also suggested by (Kaikkonen et al., 2005).

5.2 Simulations In The Laboratory

To simulate the impact of mobility and minimize learnability on the tasks, we changed obstacles’ positions in the laboratory and found this to be useful and efficient. For Task 1 (Figure 9), three of eight users in the laboratory were affected by walking in a new environment with obstacles and only one of the eight field users found walking in the lounge area a distraction to his attention to the task. For Task 2, users were asked to walk to a specified destination in both testing environments. Four of eight participants considered walking to a specified chair in the laboratory distracting while only two of eight users were affected by walking downstairs in the

field. Setting up a routine is similar to setting up obstacles; both situations required users to divide their attention while moving. Predefining a walking routine with obstacles placed in the new environment could be a good technique when simulating the walking scenario.

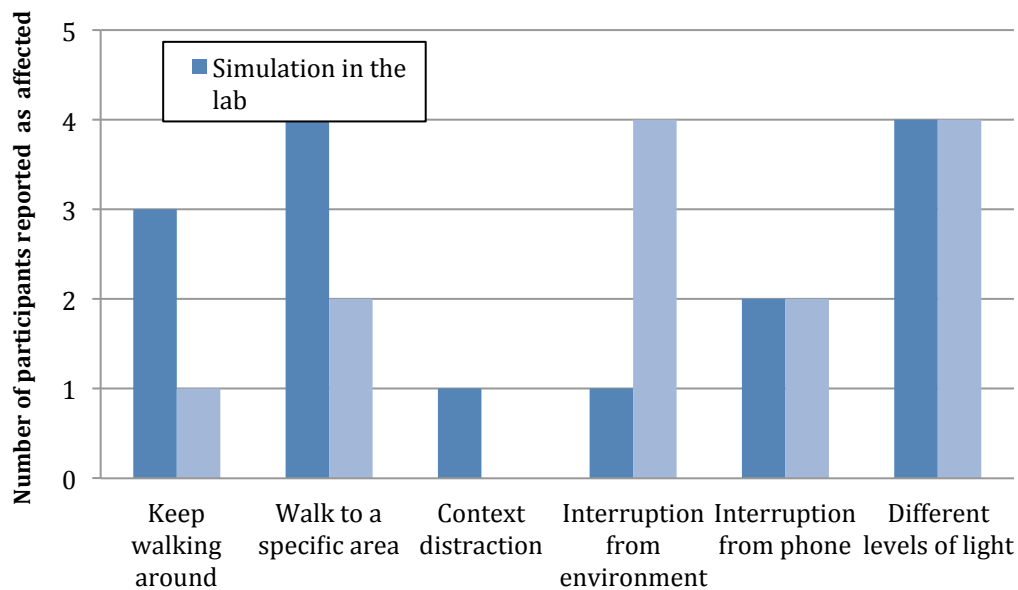


Figure 9. Users are affected by the context factors

In our study, noisy environments do not appear to affect users in the field while noises made up in the laboratory only affected one of the eight users. Betiol and Cybis (2005) also pointed out that busy and noisy environments do not affect users. This suggests that users may be used to working in a noisy complex context.

Unexpected interruptions, both internal and external, do not appear to distract users because the phone screen locks automatically during the interruptions. For Task 4, users in the laboratory were asked to stop one task to finish another task on the computer, while participants in the field were asked to stop during the task and

answer a series of simple questions. The results show that one of the eight users in the laboratory found the interruption affected their task performance, while four of the eight users in the field found the interruption problematic. This may also explain why the laboratory group participants were mindful that they were in a test while participants in the field were more easily distracted by the outside interruptions. As to the interruption from the mobile phone, a phone call was sent to the phone during Task 5. In each testing environment, two of eight users reported their performances were impacted by the unexpected phone call. One participant also considered the interruption as something he accidentally did wrong. Unexpected interruptions, both internal and external, do not appear to distract users because the phone screen locks automatically during the interruptions. We conclude that users may be used to working on their mobile systems in a noisy environment as well as with various types of interruptions.

For Task 6, the manual changing of the lights in the laboratory affected four of eight participants, and the outside natural light affected four of eight participants.

However, participants found it to be significantly more difficult to finish Task 6 in the field than in the laboratory. Changing light in the laboratory is considered annoying, but does not impede the user from completing the task. The outside natural light increases the task difficulty. To involve the light factor, an outside test is recommended.

5.3 Users' Backgrounds Affect The Results

When considering users' backgrounds, we found the results to be different between experienced users and novice users. Based on the background data collected, all of the Android users reported that the interface is similar to their own phone; so Android users were considered as experienced users in this study. We reviewed the mean execution time among the users of four different types of phones. For the first task (Task 1) during each test session, when novice users were still at the initial stage of learning, Android users took 34 seconds on average to finish the task while non-smartphone users spent 86.3 seconds on average. It is necessary to mention that for Task 5, which was considered as the most difficult task, the mean time to finish this task for Android users, Blackberry users, iPhone users and non-smartphone users were 49.5 seconds, 167.5 seconds, 135 seconds and 104 seconds respectively. Android users could quickly finish the tasks without any problems, while other novice users had to experience more difficulties to accomplish the task. Additionally, Android users' behaviors were quite different from other users' according to the observation. Two of the four Android users (one in the laboratory, one in the field) started playing around with the mobile phone when waiting for system feedback. Three of the four solved the most difficult task (Task 5) using an easier method rather than a predefined solution. All of them felt confident and satisfied during the entire test session, and could easily finish the task while being distracted. We suggest that it might be appropriate to divide users into different groups by experience because of their different behaviors. Similarly, Nielsen et al. (2006) also points out conducting a mobile evaluation experiment with only expert users would be interesting.

5.4 Recording Tool (Shootme Application) Evaluation

Our study also evaluated the new recording application ShootMe that we introduced for mobile system evaluations. This recording tool definitely has enabled users' natural interactions; users did not know the interface was being recorded during the test session through an embedded application (they did agree to record the test session in the consent form, they were only surprised at this recording method). Moreover, this recording tool directly recorded the mobile screen, so we have collected accurate and interpretable data through this method.

However, this embedded application only collected the data from touching the main screen, and it lost the data such as clicking physical button on the mobile phone. Additionally, it did not support voice recording, so a voice recorder that can gather voice comments was used in this test session. And for further data analysis, a combination of video and audio was necessary, which was time-consuming. Lastly, this installed application sometimes quitted unexpectedly because of an unknown error and it might lead to lost data.

Overall, this method of a recording application embedded in the mobile system is useful in mobile device evaluations, especially for field tests. It involves easy installation and easy data interpretation. And with the rapid development of the mobile applications, some applications that can be used to both record the mobile screen interface and collect users' comments are being developed, which can be used as recording tools for further mobile evaluations.

6 LIMITATION

There were several limitations to our study.

Firstly, with respect to participants - all of the participants were college students.

An older and more diverse demographic set of users should be recruited. And the users could be divided into different groups. It might be interesting to compare the users' behaviors between experienced users and novice users.

Secondly, there were only two users within each phone type in each testing environment. A bigger sample size is necessary to employ a more rigorous statistics analysis, especially when comparing the results within different phone types.

Thirdly, context factors were measured as categorical (affect, not affect and not sure) in this study. A rating scale of how the contexts affect users' performances would provide for stronger data analysis, which would contribute more to how to effectively simulate natural contexts in a laboratory environment.

Lastly, evaluating specific tasks rather than a general mobile system may provide more specific insights. And users' activities could focus on the task details.

7 CONCLUSION

7.1 Contributions

This study contributes significantly to mobile device evaluation exploration. Firstly, we compared the results between laboratory and field testing environment from effectiveness, efficiency, ease to use and usability problems, and further concluded

that it is insufficient to simply suggest whether field tests are necessary or not for mobile system evaluation, based on the study goals, different methods and combinations should be used. If the study is focusing on mobile system design details and usability problems identifications, a controlled laboratory environment, by recreating the natural context, is good enough to conduct the usability test, while if the study focuses on examining users' behaviors and context-dependent information, a natural field environment is recommended for providing a solid result.

Secondly, this study analyzed how to effectively simulate the natural factors in the laboratory. It seems that users are used to working on their mobile systems in a noisy environment as well as various types of interruptions, so noisy background and unexpected interruptions do not potentially affect users' performance, which suggests that it is not necessary to simulate the noisy context and the interruptions in a quiet laboratory environment. However, a walking scenario is considered as a contextual situational factor that distracts users' attention. Users need to divide their attention between the use of the mobile system and the walking situation. It is important to simulate the mobility by placing and changing obstacles in the laboratory.

Finally, this study concluded that users' background affect their performance. Users that were familiar with this mobile system were different from those who were not, especially the non-smartphone users. It is suggested that conducting a mobile evaluation experiment with only expert users would be interesting.

7.2 Lessons learned

Several lessons were learned from this study about conducting usability tests, both in the laboratory and in the field.

During the study preparation, a solid screener to recruit appropriate participants is important. Setting up a study goal and a problem hypothesis can help find the desired audience. All the test materials, such as consent form, questionnaires and detailed tasks should be prepared and consistent for all the participants. The biggest difficulty for this study was the selection of a recording tool to record the phone system interface. To enable the users' natural interactions, an embedded software needs to be used. Different phone types have different limitations, and different software require different prerequisites. To select an appropriate recording tool for Android system, the phone needs to be rooted first, which took a long time to figure it out. As a result, to prepare a mobile phone usability study, you need to make everything ready such as the test materials, the participants, the recording tools, and the mobile systems.

A pilot test is very important, especially for an uncontrolled field testing environment. For example, you need to schedule the test session in a specific time like lunchtime if you want to involve a noisy environment. During the pilot test, you might learn how to proceed in the test session when something unexpected happens. A pilot test is necessary, especially for recording tools that are used. In the pilot test, the recording software chosen (ShootMe) only allows you to record half an hour at one time, and when it exceeds half an hour, you might lose all of the recording data.

During the test session, the moderator should take some notes in case you lose the data since it is insufficient to only rely on the recording tool, either performance data or preference data. An interview is necessary after each task for confusion clarifications and the mistake corrections. To make sure you collect interpretable and efficient data is the most important thing during the test session, especially when the environment is a relaxed natural field.

8 REFERENCES

- Balagtas-Fernandez,F.,& Hussmann,H.(2009). A methodology and framework to simplify usability analysis of mobile applications. *ASE '09 Proceedings of the 2009 IEEE/ACM International Conference on Automated Software Engineering* (pp.520-524). IEEE Computer Society Washington, DC.
- Barnard,L.,Yi,J.S.,Jacko,J.A.,&Sears,A.(2005).An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices. . *International Journals of Human-Computer Studies*, 62(2005), 487-520.
- Betiol,A.Z.,&Cybis,W.A.(2005). Usability testing of mobile devices: a comparison of three approaches. *Lecture Notes in Computer Science*, 3585, 470-481.
- Beck, E. T.,Christiansen, M. K.,Kjeldskov, J., Kolbe, N., & Stage, J. (2003). Experimental evaluation of techniques for usability testing of mobile systems in a laboratory setting. *Proceedings of OZCHI 2003*, Nov 26-28, Brisbane, Australia.

- Coursaris, C. K., & Kim, D. J. (2011). A Meta-Analytical Review of Empirical Mobile Usability Studies. *Journal of Usability Studies*, 6(3), 117-171. Usability Professionals' Association.
- Coursaris, C., & Kim, D. (2005). A qualitative review of empirical mobile usability studies. *Proceedings of the 12th Americas Conference on Information Systems*, AMCIS 2006, August 04-06, Acapulco, Mexico.
- Duh, H. B., Tan, G. C., & Chen, V. H. (2006). Usability evaluation for mobile device: a comparison of laboratory and field tests. *Proceedings of the 8th international conference on Human-computer interaction with mobile devices and services*, Mobile HCI'06, September 12-15, Helsinki, Finland.
- Dumas, J. S., & Redish, J. C. (1999). *A practical guide to usability testing*. Exeter, Great Britain: Intellect Books.
- Hong, J. I., Heer, J., Waterson, S., & Landay, J. A. (2001, July). WebQuilt: a proxy-based approach to remote web usability testing. *Journal ACM Transactions on Information Systems (TOIS)*, 19(3), 263-285.
- Jensen, K. L. (2007). New methods for conducting experiments with mobile systems and services in the field. *Proceedings of the 9th international conference on Human-computer interaction with mobile devices and services*, Mobile HCI '07, September 9-12, Singapore.
- Johnson, P. (1998). Usability and mobility: interactions on the move. First Workshop on Human-Computer Interaction with Mobile Devices. Glasgow, UK 1998.

Kaikkonen, A., Kallio, T., Kekäläinen, A., Kankainen, A. & Cankar, M. (2005, November).

Usability testing of mobile applications: a comparison between laboratory and field testing. *Journal of Usability Studies*, 1(1), 4-16.

Kjeldskov, J., Skov, M.B., Als, B.S., & Høegh, R.T. (2004). Is it worth the hassle? exploring the added value of evaluating the usability of context-aware mobile systems in the field. *Proceedings of the 6th International Mobile HCI 2004 conference* (pp.61-73). LNCS, Springer-Verlag.

Kjeldskov, J., & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *International Journals of Human-Computer Studies*, 60, 599-620.

Kjeldskov, J., & Graham, C. (2003). A review of mobile HCI research methods. *Proceedings of the 5th International Mobile HCI 2003 conference* (pp.317-335).

Kjeldskov, J., Graham, C., Pedell, S., Vetere F., Howard S., Balbo S. & Davies J. (2005). Evaluating the usability of a mobile guide: the influence of location, participants and resources. *Behavior & information Technology*, 24(1), 51-65.

Lee, K. B., & Grice, R. A. (2004). Developing a new usability testing method for mobile devices. *International Professional Communication Conference 2004 IPCC 2004 Proceedings*, 115-127. IEEE.

Liisa, K., & Katja, I. (2010). Experience from the use of an eye-tracking systems in the wild. *Proceedings of the 1st International workshop on observing the mobile user experience* (pp.49-51). October 17th, 2010.

Molich, R. (2000). *Usable Web Design*. Ingeni ren/b ger, Danish.

- Magnusson,C.,Rassmus-Grohn,K.,&Szymczak,D.(2010).Methods for understanding the mobile user experience. *Proceedings of the 1st International workshop on observing the mobile user experience* (pp.1-3).October 17th,2010.
- Nauman, M., Khan, S., Zhang, X. and Seifert, J.-pierre.(2010). Beyond kernel-level integrity measurement: enabling remote attestation for the android platform. In *Proc. TRUST 2010*, Springer-Verlag, 1-15.
- Nielsen,C.M.,Overgaard,M.,Pedersen,M.B., Stage,J.,& Stenild,S.(2006). It's worth the hassle!: the added value of evaluating the usability of mobile systems in the field. *Proceedings of the 4th Nordic conference on Human-computer interaction*. NordiCHI '06: Changing Roles, October14-18, Oslo, Norway.
- Nielsen, J. (1993). *Usability Engineering*, Morgan Kaufmann Publishers, San Francisco, 1993.
- Nielsen, J. & Landauer, T. K. (1993) "A mathematical model of the finding of usability problems," *Proceedings of ACM INTERCHI'93 Conference* (Amsterdam, The Netherlands, 24-29 April 1993), pp. 206-213.
- Pascoe,J.,Ryan,N.,&Morse,D.(2000). Using while moving: HCI issues in fieldwork environments. *ACM Transactions on Computer-Human Interaction*,7(3),417-437.
- PCworld:http://www.pcworld.com/reviews/collection/3286/top_10_android_phones.html
- Rubin,J.,&Chisnell,D.(2008).*Handbook of usability testing*. New York: Wiley.

Sá,M.,& Carriço,L.(2008). Defining scenarios for mobile design and evaluation.

Proceeding of CHI '08 extended abstracts on Human factors in computing systems, CHI EA'08, April 5-10,Florence, Italy.

Schusteritsch, R., Wei, C. Y., & LaRosa, M. (2007). Towards the perfect infrastructure

for usability testing on mobile devices. *CHI 07 extended abstracts on Human factors in computing systems CHI 07*, 1839. ACM Press.

Vääätäjä, H., & Roto, V. (2010). Mobile questionnaires for user experience evaluation.

Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems CHI EA 10, 3361. ACM Press.

Waterson,S.,Landay,J.A.,& Matthews,T.(2002). In the lab and out in the wild: remote




web usability testing for mobile devices. *Proceeding of CHI '02 extended abstracts on Human factors in computing systems*, CHI EA '02, April20-25, Minneapolis, Minnesota, USA.

Participants Needed !



\$15 Walmart Gift Card is Waiting for you!

Please email to yxw5167@rit.edu ASAP!

-  A usability test about mobile phone using experience
-  A mobile phone will be provided for the test
-  A test session will last 60 minutes on campus in RIT

Mobile phone usability test Email to yxw5167@rit.edu
Mobile phone usability test Email to yxw5167@rit.edu
Mobile phone usability test Email to yxw5167@rit.edu
Mobile phone usability test Email to yxw5167@rit.edu
Mobile phone usability test Email to yxw5167@rit.edu
Mobile phone usability test Email to yxw5167@rit.edu
Mobile phone usability test Email to yxw5167@rit.edu
Mobile phone usability test Email to yxw5167@rit.edu

Appendix A2 Screener

1. What is your gender?

Male

Female

2. What is your age?

Above 35

25-34

18-25

Under 17

3. Are you using mobile phones?

Yes → go to 5 directly

No → go to 4

4. Are you going to use phone in near future?

Yes → go to 5

No → go to 8

5. What kind of phone are you using or will you use?

Smartphone → go to 6

Regular mobile phone → go to 7 directly

6. What kind of smartphone are you using or will you use?

Android system

IPhone system

Microsoft windows system

Symbian system

BlackBerry system

Others

7. How often do you use phones?

Always remember to bring it everywhere and everyday

Only use it when necessary, but do not check it quite often

Prefer other methods like emails to contact, seldom use it

8. Please leave your email address

Appendix A3 Informed consent

Rochester Institute of Technology

INFORMED CONSENT FORM

Evaluation on mobile devices in the laboratory and in the field

Principal Investigator:

You are invited to participate in a research study. The purpose of this study is to compare the evaluation results on mobile devices between the laboratory environment and the field environment. In order to enroll in this comparative study, you must give your consent to the assessment described herein. If you do not wish to participate, you may leave this study at any time.

INFORMATION

We will be assessing one type of mobile device in two different environments in this study. Our goal is not to evaluate you; we will be evaluating the mobile systems only. We are focusing on the ease of using the mobile devices in both two environments and conclude that whether a field evaluation of mobile device is necessary. During the testing session we may ask you to do some tasks using the mobile phone we provide either in the laboratory or in the field and answer some questions both relating to the tasks and some general demographic questions. The test may take approximate 1 hour.

RISKS

We do not foresee any risks associated with your participation in this research study.

BENEFITS

Participating in this study may provide the benefit of improving the research methods of collecting mobile user experience and exploring new methods in evaluating mobile devices, while gaining an inside look into how mobile devices are tested.

COMPENSATION

You will earn one \$15 Wal-Mart gift card for your time.

CONFIDENTIALITY

The information in the study records will be kept strictly confidential. Data will be stored securely and will be made available only to persons conducting the study unless you specifically give permission in writing to do otherwise. No reference will be made in oral or written reports, which could link you to the study. Publications related to this work will not make reference to any individuals. You will be given a copy of this form to keep.

CONTACT

If you have questions at any time about the study or the procedures, you may contact the researcher administrator, Yao Wang, Golisano College of Computing and

Information Sciences, Department of IST, HCI, (585) 743-7118 or e-mail
yxw5167@rit.edu.

PARTICIPATION

Your participation in this study is voluntary; you may decline to participate without penalty. If you decide to participate, you may withdraw from the study at any time without penalty and without loss of benefits to which you are otherwise entitled. If you withdraw from the study before data collection is completed your data will be returned to you or destroyed.

CONSENT

I have read and understand the above information. I have received a copy of this form. I agree to participate in this study.

Subject's signature _____ Date _____

Investigator's signature _____ Date _____

Parent's signature _____ Date _____ (if student is under 18)

Appendix A4 Orientation script

Moderator Guide

Orientation Script

Thank you very much for your interest to take part in my research study.

This is the consent form from RIT, please go ahead and read it, and if you agree, please sign it.

I am here to learn how users interact with mobile phones. This study aims to understand and improve the users' satisfaction of mobile phones. During the test session, I will ask you to use the mobile phone to finish a set of tasks. As you work on these tasks, please try to do whatever you would normally do. I will follow you to make a direct observation when you are doing the tasks.

Please remember that I am not testing you, and there are no right or wrong answers. Your doing this helps me understand what works or doesn't work about this mobile phone. This test doesn't examine you or your ability but rather how easy to use this mobile phone.

Please try to think out loud while you are working. Just tell me whatever is going through your mind when you are performing the tasks. Simply express your thoughts about what you are looking for to click and why.

Your participation is completely voluntary and you may discontinue your participation at any time. The whole session will take about 1 hour. You are welcome to take a break at any time. Please let me know if you would like a break.

Do you have any questions before we start?

Background Questionnaire

1. What kind of phone are you using?

Google's Android

Apple's IOS

Microsoft Windows Phone 7

Nokia's Symbian

RIM's BlackBerry OS

Non-Smartphone

Others

2. How often do you use phones?

Always remember to bring it everywhere and everyday

Only use it when necessary, but do not check it quite often

Prefer other methods like emails to contact, seldom use it

3. Rank the frequency of each of the following functions from 1 to 5, with 5 being used most infrequently and 1 being used most frequently.

Calling people and sending messages

Utilities like clock, notes and cameras

Receiving and replying emails

Surfing online to check weather and news

Playing games or listening to music for fun, like a handheld game console

4. How do you feel your current mobile using experience?

_Wonderful

_Moderate

_Awful

5. Is there anything of your current phone that you don't like? If Yes, Please specify.

6. Is there anything of your current phone that you like very much? If Yes, Please specify.

Start the ShootMe application on the mobile phone and start recording.

Now I'll show you the interface of this mobile device, please feel free to try it and fill in this questionnaire.

Pre-test Questionnaire

1. Does the interface look very different from your own phone?

Yes

No

I am not sure.

2. Does the interface look easy to use?

Yes

No

I am not sure.

3. Do you understand all the icons on the touch screen? If no, Please specify which.

Yes.

No _____

4. Do you understand the three buttons on the bottom? If no, Please specify which.

Yes.

No _____

5. Is the application terminology easy to understand? If no, Please specify which.

Yes.

No _____

6. Is there anything impressed you after your first look? If yes, Please specify which.

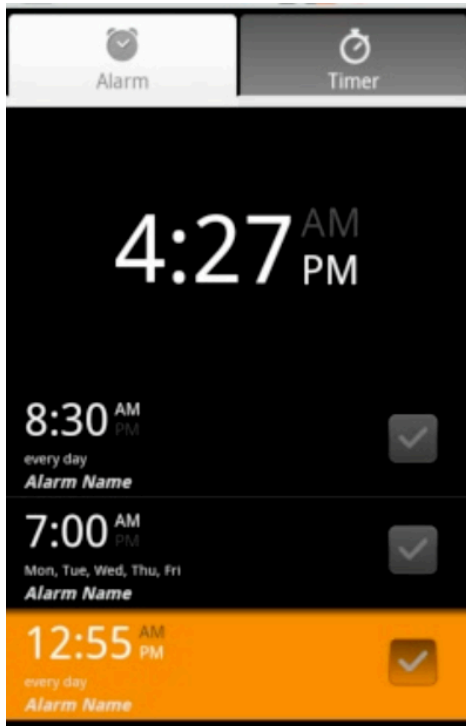
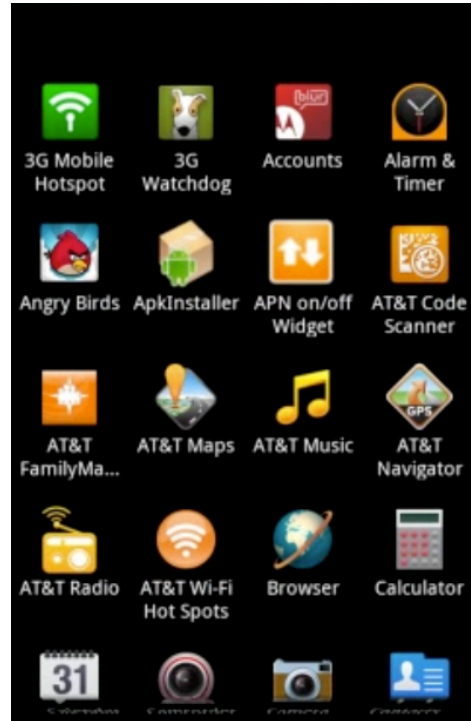
Yes _____

No.

Exercise Task Instructions

Task: Go to the alarm& timer app and delete the last alarm.

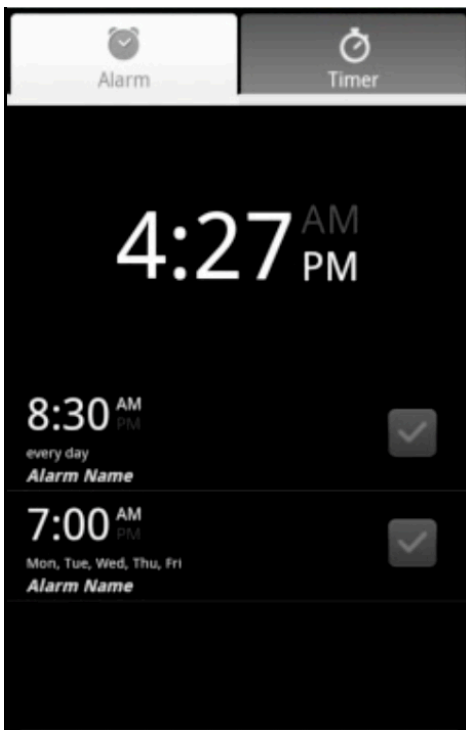
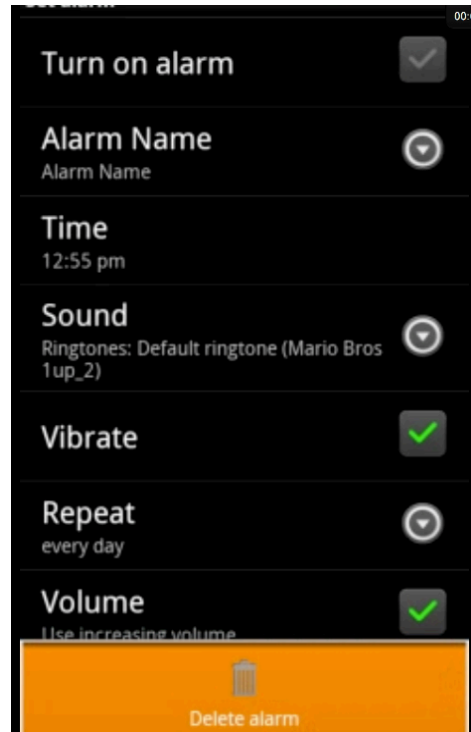
Step1: Go to the application list. Click on the alarm and timer application.



Step2: After the application launched, you may see a list of current alarms. Click the last one.

Step3: After setting alarms page loaded, Click the button on the bottom left.

A pop-up window will show indicating the function “delete alarm”.



Step4: Go back to the alarm list page, it should be only 2 alarms left.

Task Scenario

Task: Install the application

You have been waiting for the Facebook app released on android system for a long time. And now the Facebook free app is finally released in android market. You want to install it so that next time you can access to the Facebook directly by launching the application. Now I'd like you to find the "Facebook for android" app from android market and install it while you are walking around this room.

Post--Task Question:

1. How easy was it for you to find the market application?

Very Difficult

Very Easy

1 2 3 4 5 6 7

2. How easy was it for you to find the "Facebook "app in the android market?

Very Difficult

Very Easy

1 2 3 4 5 6 7

3. Do you think the installation was successful?

Yes.

No.

I am not sure.

4. Did the condition moving between chairs affect your performance?

Yes.

No.

I am not sure.

Task: Send message

You are walking around the study room, and you are thinking that you have an appointment with your friend Bob later today. But you haven't told him the time, so now you would like to send a message to Bob and tell him "See you at 7 pm" while you are moving from one chair to another chair. Please don't click send when you are finished, just leave the interface like that.

Post--Task Question:

1. How easy was it for you to complete this task?

Very Difficult

Very Easy

1 2 3 4 5 6 7

2. How satisfied were you with using this mobile phone?

Not at all satisfied

Very satisfied

1 2 3 4 5 6 7

3. Did the condition keep walking around distract your attention?

Yes.

No.

I am not sure.

Task: Change Telephone number

Imagine your friend Bob has changed his telephone number and told you that through Facebook, so you would like to change his phone number right now. Go to find the friend named “Bob”, and change his telephone number to 585-729-2735.

[During the task, a famous song “someone like you” is playing.]

Post--Task Question:

1. How easy was it for you to edit the telephone number?

Very Difficult

Very Easy

1 2 3 4 5 6 7

2. Do you think the phone number was changed successfully?

Yes.

No.

I am not sure.

3. Did the song distract your attention?

Yes.

No.

I am not sure.

4. Was the edit button where you expected to find it? If no, please write your expectations.

Task: Reply email

You are checking the email using your mobile phone. And you find one email sent from Yao Wang, which you didn't have a chance to reply. So please reply the email using your mobile phone now. The body will be "hello, the weather outside is pretty good". **[During the task, you would be interrupted and asked to search your favorite song. When you finished, go back and continue your task.]**

Post--Task Question:

1. How easy was it for you to use the email function of this mobile phone?
Very Difficult Very Easy
1 2 3 4 5 6 7
2. Do you think the email was sent out successfully?
_Yes.
_No.
_I am not sure.
3. Was the reply button where you expected to find it? If no, please write your expectations.
4. Did the interruption task distract your attention?
_Yes.
_No.
_I am not sure.
5. What do you think of the virtual keyboard on this mobile phone?

Task: Uninstall application

Imagine that most of your friends no longer use “Facebook” app, so you decide to delete it since you only use it for social activities. Now I’d like you to go to uninstall this app. **[During the task, an unexpected call will reach your mobile phone.]**

Post--Task Question:

1. How easy was it for you to uninstall this app?

Very Difficult

Very Easy

1 2 3 4 5 6 7

2. Do you think you uninstall the app successfully?

Yes.

No.

I am not sure.

3. Did the unexpected phone call affect your performance?

Yes.

No.

I am not sure.

4. What were you thinking when the unexpected call was reaching?

5. Was the uninstall function where you expected to find it? If No, Please write your expectations.

Task: Search restaurants

You are playing the mobile phone in a changing light environment. You would like to go out for dinner to try some new Chinese food, so you need to find some Chinese restaurants near you. Go to maps and search Chinese restaurants near you. **[The light will keep changing during the task.]**

Post--Task Question:

1. How easy was it for you to complete this task?

Very Difficult

Very Easy

1 2 3 4 5 6 7

2. How satisfied were you with using this mobile phone?

Not at all satisfied

Very satisfied

1 2 3 4 5 6 7

3. Did the changing light environment affect your performance?

Yes.

No.

I am not sure.

4. Was the LCD brightness good enough for a changing light environment?

Yes.

No.

I am not sure.

Post-test Questionnaire

1. Do you think this mobile phone was easy to learn?

Yes.

No.

I am not sure.

2. How well was the application list organized?

Very Bad

Very Well

1 2 3 4 5 6 7

3. How clear or understandable was the navigation?

Very Confused

Very Clear

1 2 3 4 5 6 7

4. How clear or understandable were the icons in the application list?

Very Confused

Very Clear

1 2 3 4 5 6 7

5. How easy was it for you to use touch screen to finish the tasks?

Very Difficult

Very Easy

1 2 3 4 5 6 7

6. How easy was it for you to use bottom three buttons to finish the tasks?

Very Difficult

Very Easy

1 2 3 4 5 6 7

7. Is there anything that you didn't like about this mobile phone? If yes, Please specify.

8. Is there anything that you liked best about this mobile phone? If yes, Please specify.

Debriefing

Review the test session.

Do you have any questions?

I appreciate your participation. The information and data I collect today from you will be only used for research analysis. Your performance will help me to improve the mobile phone design. If you have any further questions or concerns, you are very welcome to contact me at any time.

Thank you again for you taking time to participate.

Give the gift card.

Appendix A5 Usability problems identified in the laboratory

Usability problems	P1	P2	P3	P4	P5	P6	P7	P8
1. Market application name is confused					X			X
2. Some terminologies are confused (e.g., settings, preferences)	X			X			X	X
3. Respond inline (email application) is not interpretable								X
4. The application icon is not understandable		X						
5. Menu button is not understandable	X	X	X	X		X	X	X
6. Reply button (email application) is ignored								X
7. Send icon is hard to understand					X		X	X
8. Search icon (market application) is hard to understand						X		
9. Trash icon on the desktop is misleading								X
10. Check icon (Alarm application) is misleading						X	X	X
11. Phone call icon is ignored			X		X		X	X
12. Reply icon (email application) is similar to back button							X	
13. Message icon is similar to email icon				X			X	
14. Search icon (market application) is too small							X	
15. Virtual keyboard is too small				X	X	X	X	
16. Virtual keyboard is too sensitive				X			X	

17.Hard to switch between words and numbers				X			X	
18.Edit function is hard to find	X			X	X			
19.Hard to find the desired application from the application list		X		X	X		X	X
20. Do not know how to unlock the screen	X	X			X		X	
21. Do not know what to input in the search box	X			X	X			
22.Do not know how to find the applications		X						
23.Do not know how to uninstall the application	X	X			X	X		X
24.Edit function is not related to the application	X		X	X	X	X	X	X
25.Applications are not grouped based on the functions		X					X	
26.Menu function is hard to understand		X		x	X		X	
27.Hard to switch between the main screen and bottom buttons		X						X
28.The function buttons are not available on the main screen			X	X	X	X	X	
29.The uninstall process is not simple or mobile	X	X		X	X			X
30.Uninstall feedback is not appropriate		X						X
Total: 30	8	11	5	13	12	7	17	15

Appendix A6 Usability problems identified in the field

Usability problems	P1	P2	P3	P4	P5	P6	P7	P8
1. Market application name is confused		X						
2. Some terminologies are confused (e.g., settings, preferences)		X			X			
3. Respond inline (email application) is not interpretable							X	X
4. The application icon is not understandable					X			X
5. Menu button is not understandable					X	X	X	X
6. Send icon is hard to understand							X	X
7. Trash icon on the desktop is misleading					X			
8. Check icon (Alarm application) is misleading	X							
9. Phone call icon is ignored					X			
10. Reply icon (email application) is similar to back button							X	X
11. Search icon (market application) is too small					X		X	
12. Virtual keyboard is too small						X		
13. Virtual keyboard is too sensitive			X					X
14. Hard to switch between words and numbers		X						
15. Hard to find the desired application from the application list						X		X
16. Hard to see the screen in the outside weather			X	X				
17. Do not know how to unlock the screen						X		

18. Do not know what to input in the search box							X	
19. Do not know how to uninstall the application	X	X	X		X	X	X	X
20. Edit function is not related to the application			X		X	X		
21. Applications are not grouped based on the functions			X		X	X		
22. Hard to switch between the main screen and bottom buttons		X						
23. The function buttons are not available on the main screen			X			X	X	
24. The uninstall process is not simple or mobile	X							
25. WIFI connection feedback is not clear			X		X			
26. Uninstall feedback is not appropriate	X						X	
Total: 26	4	5	7	1	10	8	9	8