MPhil Thesis

Comparison of Field and Laboratory Usability Testing for Mobile Text Entry

Submitted By: Sohail Azad

Supervised by: Dr Mark D Dunlop

Computer and Information Sciences Department
University of Strathclyde Glasgow

Submitted in partial fulfilment of the requirements for MPhil in the Department of Computer and Information Sciences at The University of Strathclyde

Signed: Date:

Table of Contents

Abstract

Mobile phones have gained considerable popularity and are now part of our daily life and are being used for diverse purposes other than making and receiving phone calls. It has become important from a user's point of view that the product they get is more user friendly. At present most ~~of~~ usability evaluations have been done in a lab, where we have a controlled and peaceful environment but this does not represent the true environment and conditions in which mobile phones are being used. This study was designed to look into the work done so far in usability evaluation techniques for lab, field and the comparison~~s~~ of the two to get an idea of what other researchers have to say. Then conduct a usability evaluation in the lab and in field with text entry as the main activity to see what conclusion can be drawn.

Based on data gathered during this study, it was noted that users are slower in the field as compared to the lab and there were significant differences in the results of lab and field usability evaluations. Another important factor discovered in the second study was the difference in performance of native and non-native English speaker's at a text entry task this study showed that there was a significant difference in performance of native and non-native English speakers. This suggested that culture also effects usability testing of mobile devices. Usability evaluation done in one culture may reveal different results if done in a different culture. This will form the basis for future studies.

Chapter 1
Motivation and introduction

1.1 Background

Looking at the history of mobile devices between 1980s and 2000s, the mobile phone has gone from being an expensive item used by the business elite to a pervasive, personal communications tool for the general population. In most countries, mobile phones outnumber land-line phones, with fixed landlines numbering 1.3 billion but mobile subscriptions 3.3 billion at the end of 2007 and the figures are still rising.

1.2 Use of Mobile Phones

As well as making voice calls, mobile phones can be used for things like:

- Sending and receiving emails, text and multimedia messages
- Registering contacts
- Using calculator, currency and alarm functions etc
- Internet
- Playing games
- Taking photos and videos

This thesis will focus on text entry on mobile phone only. Initial growth of text messaging was slow, with customers in 1995 sending on average only 0.4 messages per GSM customer per month [1]. If we look around today we know that SMS text messaging is the most widely used mobile phone data application on the planet, with 2.4 billion active users, or 74% of all mobile phone subscribers sending and receiving text messages on their phones [2].

Motivation

As use of mobile devices is growing and is being used in diverse cultures and environments it became important to develop mobile devices and applications which are user friendly and address diversity. Globally, just over one-fourth (28%) of mobile phone owners worldwide have browsed the Internet on a wireless handset according to *The Face of the Web* study. SMS text messaging remains the most popular activity among consumers, while other communication-based wireless activities are also growing. Over half (52%) of all mobile phone households today have sent or received a text message, and over a third (37%) have sent or received e-mail on a mobile phone [3]. This growing trend in use of mobile phones other than talking to people encouraged me to look into the usability side of the text entry and find out what research has been done so far and how it can be improved.

Lab vs. Field Usability

As most mobile phone usability evaluations are being done in the lab which is not an ideal environment but testing outside the lab has some drawbacks. The environment outside of the lab is often changing rapidly without forewarning, is difficult or impossible to control, and has other environmental and operational constraints that cannot easily be modelled in a laboratory, partly because the necessary level of ecological validity is almost impossible to achieve in the artificial lab environment. Field study evaluation techniques are insufficient in these environments. Furthermore, it is often difficult or impossible to ascertain which behavioural data is needed to answer questions about user requirements, UI design, and user acceptance. [4]

This study focused mainly on the comparison of field and laboratory usability testing for mobile text entry. This was a pilot study with the aim to provide a basis to conduct comprehensive experiments in the future to identify best possible methods to conduct ~~a~~ usability testing in the lab that depict results which are as close as possible to real world data.

All phases of this study were conducted in the lab and in the field. A program written in J2ME was used to let users enter text. All inputs by users were recorded on the memory card of the mobile phone. Data collected during study is valuable because it is important to establish an idea of how environment effects text entry capability of users. One interesting observation that came to light during this study was the difference in performance of native and non-native English speakers, it was decided to conduct a second study to get a clearer picture about effects of language on usability evaluations.

Effect of Language on Usability

This study was based on the hypothesis that people from different cultures perform differently in text entry tasks. This could be due to language limitations or other factors such as education, social structure, religion etc.

This study was designed on similar lines to the first study. All the phases of the experiment were conducted in the lab and in the field. The program was written in J2ME and was used to let users enter text and all inputs were recorded on the memory card. Participants were divided into two groups based on their native language. The first group had English as their mother tongue and the second group had a language other than English as their mother tongue. This study was designed in a way to minimise the effects of the non-native English group having problems with spelling~~s~~. This will be explained in the detail in later chapter.

Chapter 2

Literature Review

This chapter will introduce you to the term usability testing and describe why it is important and will also explain work done in the domain of text input in mobile phones. This chapter will also cover research carried out to compare field and lab usability testing and then help us understand which approach is suitable for mobile devices and how we can make improvements based on past work.

2.1 Usability

Generally people define usability as “Efficiency with which a user can perform required tasks with a product”. [5]

The document ISO 9241-11 (1998) guidance on usability, issued by the International Organization for Standardization, defines usability as: “The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.”

In simple words usability is a term used to denote the ease with which people can use a particular tool or other human-made object in order to achieve a particular goal. Usability can also refer to the methods of measuring usability and the study of the principles behind an object's perceived efficiency or elegance.

Usability consultant Jakob Nielson and computer science professor Ben Shneiderman have written (separately)[6][7] about a framework of system acceptability, where usability is a part of "usefulness" and is composed of:

- **Learnability:** How easy is it for users to accomplish basic tasks the first time they encounter the design?
- Efficiency: Once users have learned the design, how quickly can they perform tasks?
- **Memorability:** When users return to the design after a period of not using it, how easily can they re establish proficiency?
- **Errors:** How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
- **Satisfaction:** How pleasant is it to use the design?

2.2 Usability testing

Usability testing is a technique used to evaluate a product by testing it on users. This can be seen as an irreplaceable usability practice, since it gives direct input on how real users use the system.

2.3 Usability testing techniques

In usability testing approach, some of the techniques mentioned by J. Nielsen are as follows [6]

a. Coaching Method

This technique can be used for usability test, where the participants are allowed to ask any system-related questions of an expert coach who will answer to the best of his or her ability. Usually the tester serves as the coach. One variant of the method involves a separate expert user serving as the coach, while the tester observes both the interaction between the participant and the computer, and the interaction between the participant and the coach.

b. Co- Discovery Learning

During a usability test, two test users attempt to perform tasks together while being observed. They are to help each other in the same manner as they would if they were working together to accomplish a common goal using the product. They are encouraged to explain what they are thinking about while working on the tasks. Compared to thinking-aloud protocol, this technique makes it more natural for the test users to verbalize their thoughts during the test

c. Performance Measurement

This technique is used to obtain quantitative data about test participants' performance when they perform the tasks during usability test. This will generally prohibit an interaction between the participant and the tester during the test that will affect the quantitative performance data. It should be conducted in a formal usability laboratory so that the data can be collected accurately and possible unexpected interference is minimized. Quantitative data is most useful in doing comparative testing, or testing against predefined benchmarks. To obtain dependable results, at least 5 user participants are needed, while 8 or more participants would be more desirable. The technique can be used in combination with retrospective testing, post-test interview or questionnaires so that both quantitative and qualitative data are obtained.

d. Question Asking Protocol

During a usability test, besides letting the test users to verbalize their thoughts, as in the thinking aloud protocol, the testers prompt them by asking direct questions about the product, in order to understand their mental model of the system and the tasks, and where they have trouble in understanding and using the system. This is a more natural way than the thinking-aloud method in letting the test user to verbalize their thoughts.

e. Remote Testing

Remote usability testing is used when tester(s) are separated in space and/or time from the participants. This means that the tester(s) cannot observe the testing process directly and that the participants are usually not in a formal usability laboratory. There are different types of remote testing. One is same-time but different-place, where the tester can observe the test user's screen through the computer network, and may be able to hear what the test user says during the test through speaker telephone. Another is different-time different-place testing, where the user's test session is guided and logged through a special piece of software as well as additional code added to the system being tested.

f. Retrospective Testing

If a videotape has been made of a usability test session, the tester(s) can collect more information by reviewing the videotape together with the user participants and asking them questions regarding their behaviour during the test. So this technique should be used along with other techniques, especially those where the interaction between the testers and the participants is restricted. But using this technique means that each test takes at least twice as long. Another obvious requirement for using this technique is that the user's interaction with the computer needs to be recorded and replayed.

g. Shadowing Technique

During a usability test, the tester has an expert user (in the task domain) sat next to him/her and explains their behaviour to the tester. This technique is used when it's not appropriate for the test user to think aloud or talk to the tester while working on the tasks.

h. Teaching Method

During a usability test, let the test users interact with the system first, so that they get familiar with it and acquire some expertise in accomplishing tasks using the system. Then introduce a novice user to each test user. The novice user~~s~~ is briefed by the tester to limit their active participation and not to become an active problem-solver. Each test user is asked to explain to the novice how the system works and demonstrate to him/her a set of pre-determined tasks.

i. Thinking aloud Protocol

During the course of a usability test, the test users are asked to verbalize their thoughts, feelings, and opinions while interacting with the system. It is very useful in capturing a wide range of cognitive activities.

2.4 Why usability is important

A highly usable system offers benefits to both users and business. The primary benefit to users is that they are able to achieve their tasks easily and efficiently. This sounds simple, but the feeling of achievement that people get when they use a computer system without frustration should not be underestimated [8]. Success and failure of products can be measured with this simple rule of usability.

We can summarise usability benefits in the following points

It

- Reduces development time and costs
- Reduces support costs
- Reduces user errors
- Reduces training time and costs
- Provides a return on Investment.

Since this study will address usability testing of mobile devices. Let us familiarise our self with mobile computing.

2.5 Mobile Computing

Mobile computing is referred as ability to use technology in remote or mobile environments [9]. In terms of mobile computing, devices that come to mind are

- Mobile Phones
- Smart Phones
- PDAs

Mobile devices are being used by people on the move or in environments where their level of concentration is different ~~as~~ compared to an office or room. For this reason it becomes important to have a product that is user friendly and takes into account the environment~~s~~ it will be used in and the diversity of users. Mobile phones being the most popular of mobile devices available give us more freedom and are easy to carry.

2.6 Mobile Usability

If we look at the mobile phone market we will find that at present there are more than 2 billion and according to an estimate by GSMA that new users are signing at the rate of 1000 per minute in the world [10]. New markets like China, India, Africa and Latin America are playing a major role in the boost of the mobile market. All companies are trying their level best to grab the major portion of the market. They need to introduce products which catch people's attention and in order to do this they must concentrate on usability. A good product from a usability point of view will attract the attention of many people. Naturally that will improve the sales and will attract people who find it difficult to interact with mobile phones.

2.7 Text Entry in Mobile Devices

Text entry in mobile devices started with sms and now mobile phones are being used to enter text in almost all the applications which we use on normal PC. Now features like use of internet on mobile phones, mobile mail, notes, to do lists, different versions of Word, Excel, Power Point, and Adobe PDF have transformed the way mobile phones are used. With this transformation it has become very important to look into ways which will make text entry on mobile phones more easy and popular, especially when talking about a product which is being used all around the world with diverse languages, because in a global context, text entry is far from simple. Thousands of languages and hundreds of writing systems make creating a unified system for text entry a daunting task. Same is true for modelling human text processing and writing. A global market for electronic devices means manufacturers of text processing systems must acknowledge and meet the needs of a diverse user community. [12]

Users currently use methods such as multi-tap or Tegic Communications' 12-key phone to enter text using a numeric keypad. Novice users of these methods achieve text entry rates of 5-10 words per minute and experts around 20 [13].

2.8 Traditional approach to text input

Traditional approach to enter text on mobile phone is through multi-tap. This is most commonly used approach by mobile phone users all around the world [14], users have to press a button multiple times to get the required character (e.g. in order to type c users will press 222). Here inputting text may be time consuming especially when users have to wait for time out between subsequent letters on the same button. [15].

2.9 Predictive text Input

Predictive text input system was introduced to overcome problems in multi tap. Instead of pressing single key many times users have to press each key once and that key is then mapped to most commonly used word in that combination (e.g. 8447 will mean *This*). Although one combination of keys can be mapped to more than one words but solutions are available and in practice to overcome this problem [15]

2.10 Multi-tap vs. Predictive

Predictive text is gaining more popularity than multi-tab and as predictive text entry mapping techniques improve with time it will become most widely used text input system in world. Experiments done to compare multi tab with predictive have proven that predictive out performs multi-tap with text entry speed rising from 8wpm in case of multi-tap, to 20 wpm for predictive(T9)[15]

2.11 Touch screen key boards

Many high end mobile phones like Nokia 5800 and Apple iPhone have started using touch screen as an input and interaction mode. This trend has resulted in keypads being removed in most cases and leaving all area for mobile screens. Now users have access to full QWERTY key boards on screen, users can now enter text using their fingers or stylus. It has brought key entry in mobile devices closer to normal personal computers. In a recent study expert iPhone users were able to achieve 60wpm [15]

2.12 Future Trends

Future of text entry in mobile devices is not clear and this is supported by a review done in the latest research [15].There is still an ongoing debate which text entry mode will prove successful in future. One of the major factor in this regard will be the future trend in size of mobile devices, small size means less space for key pads and this may change text entry trend completely but current trends are pointing towards QWERTY being used to enter text on mobile devices [15] in most cases it will be in form of touch screen.

2.13 Challenges in Text Entry

On mobile devices the major obstacle is designing a simple and efficient interface for text entry. These mobile devices do not possess enough space to accommodate complete keyboard configurations available in a normal computer, and with a trend of reduced size the problem has become more complex [26] .A single button on the input keypad may be assigned to more than one character, thus creating plurality in character resolution that requires disambiguation schemes to identify the original character intended by the user.

Smaller and lighter phones are in more demand but smaller size comes at the expense of limited display for visual feedback and less space for interaction. [17] This shrinking size problem is also stated by other researchers [18]. This proves that there is a problem which people have been looking into for a long time.

2.14 Text input research

As a mobile devices user we spend a lot of time carrying out text input activities. This time is generally divided into two main parts - input time and correction time. The study of text input has been around for some time. The following picture will give you an idea about the work done in this field.

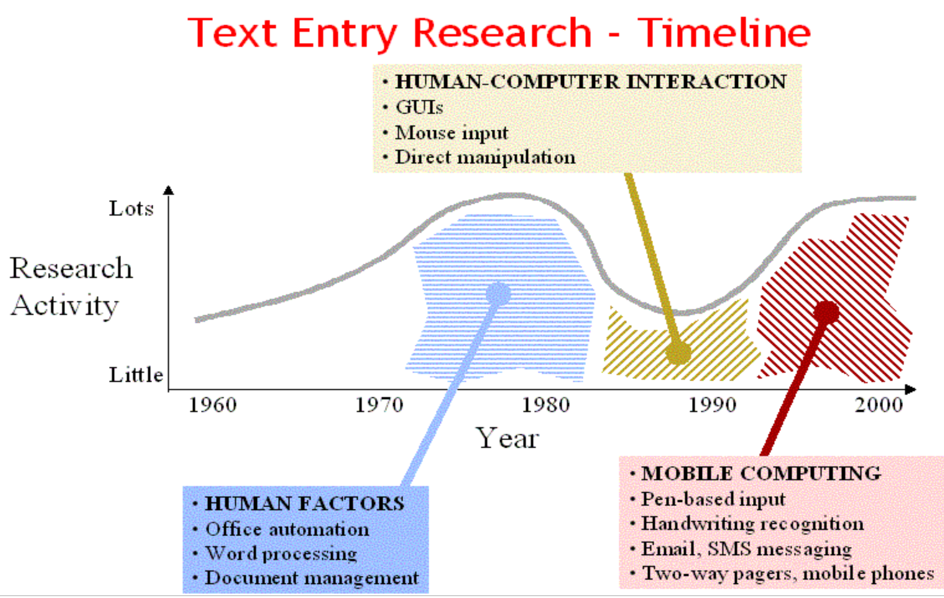


Figure 1 Text Entry Research time line [19]

Considerable work has been done by Mackenzie in this field [20][21][22][23]. Mackenzie has even proposed and published a list of text entry phrases in the hope that researchers might use this list and thereby allow for comparisons across research studies that are carried out by different groups [24]. In this experiment phrases were selected from this list.

2.15 Text entry share in new research

Looking on papers presented in the last two Mobile HCI conferences indicates the share of Text Entry Research done in mobile devices field. In Mobile HCI 2007 total papers presented were 37 out of these 5 were related to text input that equals 13.5% which was highest share of research papers presented in Mobile HCI 2007.e.g An Evaluation of Stylus-Based Text Entry Methods on Handheld Devices in Stationary and Mobile Settings by Koji Yatani and Optimized Layout for Keypad Entry System by Arpit Mittal.

In Mobile HCI 2008 papers related to text input were concentrated on the touch screen like Yong S. Park Sung H. Han Jaehyun Park and Youngseok Cho [25] looked at the effects of touch key sizes and locations on the one-handed thumb input that is popular in mobile phone interactions. Similarly Robert Hardy, Enrico Rukzio [26] presented Touch & Interact: an interaction technique in which a mobile phone is able to touch a display, at any position, to perform selections and is comparable to approached used on touch screen. This shows future trend towards touch screen interfaces for text entry systems.

2.16 Field Vs Lab

For several years now this topic has been controversial in human computer interaction and in case of usability evaluation in particular. Although most of the usability testing is conducted in laboratories around the world many people have been resisting this idea by supporting testing in the field [27].

Established concepts, methodologies, and approaches in human–computer interaction are being challenged by the increasing focus on systems for wearable, handheld, and mobile computing devices [28]. This move beyond office, home, and other stationary use settings has created a need for new approaches to design and evaluate useful and usable systems [29].

A usability evaluation of a mobile system should always be conducted in the field. It is important that systems for mobile devices are tested in realistic settings, since testing in a conventional usability laboratory is not likely to find all problems that would occur in real mobile phone usage [30].

Practically speaking there has not been much evaluation done in the field so far. Almost 61% of mobile evaluation is done in laboratory and 22% are done in field based on selection of 45 papers [31].

2.17 Restrictions in the Field

There are some restrictions in field usability testing. It is more time consuming, data collection becomes complicated, and uncontrolled variables reduce experimental control. It also becomes very difficult to observe participants. In some cases there are health and safety issues involved in conducting a usability testing in the field.

2.18 Laboratory

In laboratory it's safer to conduct experiments. The environment is much safer and it allows us to collect high quality data. [32] Usability testing in the lab is cheaper and gives more control. Users can perform their task without any disturbance and it gives observers a good opportunity to observe participants in more details and get useful data.

2.19 Confusion in lab and field evaluations

It is worth conducting tests in the field, even though it is problematic. But in the past few years contradictory papers have been published. One stating that usability evaluations in the field is very limited and recreation of central aspects of the use context in a laboratory setting enables the identification of the same usability problems [33] . Another saying that if evaluation is conducted in the same way both ~~is~~ in the lab and in the field then there is a significant difference in the results. Field evaluations are more successful as this setting enables identification of significantly more usability problems compared to the laboratory setting [34].Then another paper stated that was no difference in the number of problems that occurred in the two test settings. [35]

A paper presented in Mobile HCI 2006 Finland stated that " The analyses of the comparison between usability testing done in two different settings revealed that there were many more types and occurrences of usability problems found in the field than in the laboratory".[36]

After reviewing these papers it was felt that there is a great need to study this in more detail so that we can settle this argument. In order to achieve this it was important to be able to conduct usability tests whose results depict the problems faced by users in real world. This pilot study is just a way to look into this problem then work on it in more detail so that a proper and clear picture emerges in front of us.

2.20 Share of Evaluations studies in Research

In mobile HCI 2007 out of 37 papers 5 were related to evaluation and design. This is 13.5% of total papers presented. These 5 papers did evaluations of latest technologies like NFC, mobile business services, virtual reality, web based applications. In mobile HCI 2008 also had 3 papers based on evaluation techniques used for innovation styles and application. There was no papers presented in Mobile HCI 2007 that addressed the direct comparisons of lab and field usability evaluation.

2.21 Way Forward

In ~~a~~ the latest research , a researcher compared lab usability field usability and real life usability and observed that you find a high number of usability issues in lab but low number of usage issues in lab and in the field you see less usability issues and more user issues [37]. Some researchers propose that lab evaluations are more efficient in identifying cosmetic problems, which do not hinder interaction and user performance. Second, field evaluation is more likely to identify issues that are related to the real context of use, such as navigation and social comfort [38].

If we look at the research papers from 2007 onwards addressing text entry in mobile devices we will find that only one study was done in the field [39]. This case study was done in an underground train. In another study done by Koji Yatani and Khai N. Truong [40] users were supposed to walk on a designated path. This was a good attempt to make a lab environment as close to real world as possible.

So a way forward is a new hybrid approach in which users can perform cooperative evaluation sessions [46] in real world contexts, such as a cafe etc. and that will be the main aim for this study as well. Testing text entry of mobile devices in the lab and in the field will give an insight into how different results can be when two experiments are conducted in different locations.

Chapter 3

Pilot Studies

Based on the literature review in previous chapter it was observed that there are contradictions in findings when it comes to comparisons of field and lab usability evaluation techniques. This was kept in mind when designing the experiment for this pilot study, so that this study can help us to decide that there are differences in lab and usability evaluation results.

3.1 Equipment and Subjects

In order to conduct experiment three mobile phones were chosen. The reason for choosing these mobile phones was the best possible option for the phone that had best set of keys.

1) Nokia N70 — a traditional ISO 9995-8 12 key phone
2) Blackberry — a phone with 20-key keypad
3) Palm tungsten c — a PDA with high quality mini-qwerty key board

Nokia N70

Blackberry 7130g

Palm tungsten c

The study was designed to compare these keyboards and to see if there are any differences in lab and field usability evaluations and to verify the contradicting results presented in the papers discussed in literature review.

And for each phone a group of 10 phrases was chosen and each set of phrases consisted of equal number of characters. During experiments the order of these three sets of phrases was changed so that every user gets to enter a different set of phrases on every phone and the phone, phrases order is balanced.

3.2 Phrases for Experiment

All of the phrases in three different sets were selected from a large set proposed by Mackenzie [47].

A:-

Do not worry about this
My bike has a flat tire
I am going to a music lesson
Do you feel too bad about it
Just in time for the party
Parking lot is full of trucks
Why do you ask silly questions
Keep receipts for all your expenses
I cannot believe I ate the whole thing
I do not fully agree with you

B:-

Be home before midnight
I will meet you at noon
Do you like to go camping
Work hard to reach the summit
Assignments are due tomorrow
Where did I leave my glasses
Please keep this confidential
My favourite sport is football
Look in the syllabus for the courses
Please try to be home before midnight

C:-

Thank you for your help
He is still on our team
It is very windy today
On the way to the cottage
My bank account is overdrawn
He is just like everyone else
I took the rackets from the shop
Bank transaction is not registered

Would you like to come to my house

The elevator door appears to be stuck

All sets of phrases consisted of equal number of characters and it was allocated to users in a way that reduces repetition on more than one keypads. Phrases allocation procedure is presented in table 1 and table 2.

User Number	Phrases to Enter on 12 Key	Phrases to Enter on 14 Key	Phrases to Enter on QWERTY
1	A	B	C
2	B	C	A
3	C	A	B
4	A	B	C
5	B	C	B
6	C	A	B
7	A	B	C
8	B	C	A
9	C	A	B
10	A	B	C

Table 1:- Allocation Of Phrases for Lab Experiment

User Number	Phrases to Enter on 12 Key	Phrases to Enter on 14 Key	Phrases to Enter on QWERTY
1	C	A	B
2	A	B	C
3	B	C	A
4	C	A	B
5	A	B	C
6	B	C	A
7	C	A	B

8	A	B	C
9	B	C	A
10	C	A	B

Table 2:- Allocation of Phrases for Field Experiment

3.3 Participants

A total of 10 participants were selected to take part in this experiment all of them were volunteers from the Department of Computer and Information Sciences with experience of using mobile phones and sending text messages. No one was paid for this experiment. 25 students were contacted by e-mail and in person, out of which 13 showed interest and a final 10 selected were the ones with experience of using any of the phones being used in experiment. 5 participants were native English Speaker and 5 were non native English speakers.

3.4 Data collection

Data from the experiment was collected firstly on a registration form where data related to age, gender, and experience with mobile phones, ethnic origin and texting behaviour was noted. Data collected is mentioned in observation section of this chapter. There was logging software on each phone to keep a track of time taken to complete the task and the errors done while they were entering text. More detail on this in next section. NASA TLX forms were used to get feedback about mental and physical stress during experiments.

3.5 Text Input System

In this experiment, predictive text input system with a built in dictionary was used and in order to keep track of time taken by the participants and mistakes done by them code was modified in way that memory card of mobile phones in use during experiment can be used as a storage location. It started storing data when first key was pressed and then kept record of time taken by users between each space key pressed in milliseconds. It also logged any spelling mistakes done by users. Time logged in this experiment is the time taken between first character and last character typed.

This software was originally made to test 12 key and 5 key phones but it was changed for 14 key and QWERTY key pads using J2ME. Since it was a predictive text entry system it worked on the same principals as 12-key phone text input system.

It has a built in dictionary of the top 755 words from a language dictionary. The dictionary was kept small so that it can run on the platforms with limited memory. But it had all the words used in the phrases. Since all dictionaries are sorted by frequency the omissions of rare words would not affect the study as higher frequency words will always be suggested before rare words.

3.6 Usability Evaluation Techniques Used

Usability evaluation techniques used for this experiment were a combination of teaching method and shadowing techniques it was chosen to meet requirements of this study from a list of evaluation techniques discussed in previous chapter.

3.7 Experimental Design

10 participants took part in the lab experiment and 10 in the field. Participants taking part in the lab and the field were the same participants. The reason for keeping the participants the same was to observe the difference in performance by the same people in different locations. All participants were given a briefing about the experiment and how they were supposed to complete different tasks then they were given time to practice. The experiment then started when participants were ready. Before starting the experiment they were assigned user numbers and information about their age, experience with mobiles, and the extent to which they used mobile phones to send sms was gathered. They were told it's not a test of their typing skills but focus is on the effects of number of keys on key pad on any individual. After completion of each part of the experiment they were given NASA TLX form to get feedback.

3.8 Lab Experiment

Lab experiment was conducted in a room where they had no disturbance they performed all their tasks while sitting on the chair. They had set of phrases in front of them and were entered by looking at the page lying on the table. During this experiment users were not disturbed and allowed to do their work so that it does not affect results. The following figures are from that experiment.

Fig: 1

3.9 Field Experiment

The field experiment was designed to be a bit complex with the user having to move along a path which demanded their attention while they were moving. They had to vary speed, change directions and in some part had to go up steps or come down. As you can see from the picture below the route was such that they had to be aware of their surroundings while completing the task. It was a good way to depict a real world situation. Where they had to be aware of surroundings and perform the task in an environment where there are lots of distractions. And because of ~~a~~ building work going on nearby there was a good amount of noise that also had an effect on concentration of users.

Fig2: Location of Field Experiment

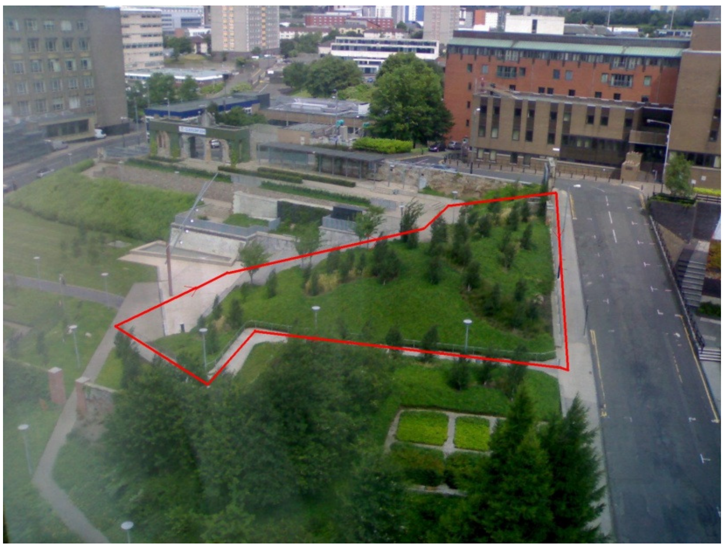


Fig 3: Route chosen for experiment

Fig 4: Test in progress

3.10 Observations

Average age of participants was 28.5 years. Based on data collected from participants the following facts emerged.

How often users used mobile phones for sending text messages?

Many times a day	60%
Once a day	10%
Once a week	30%

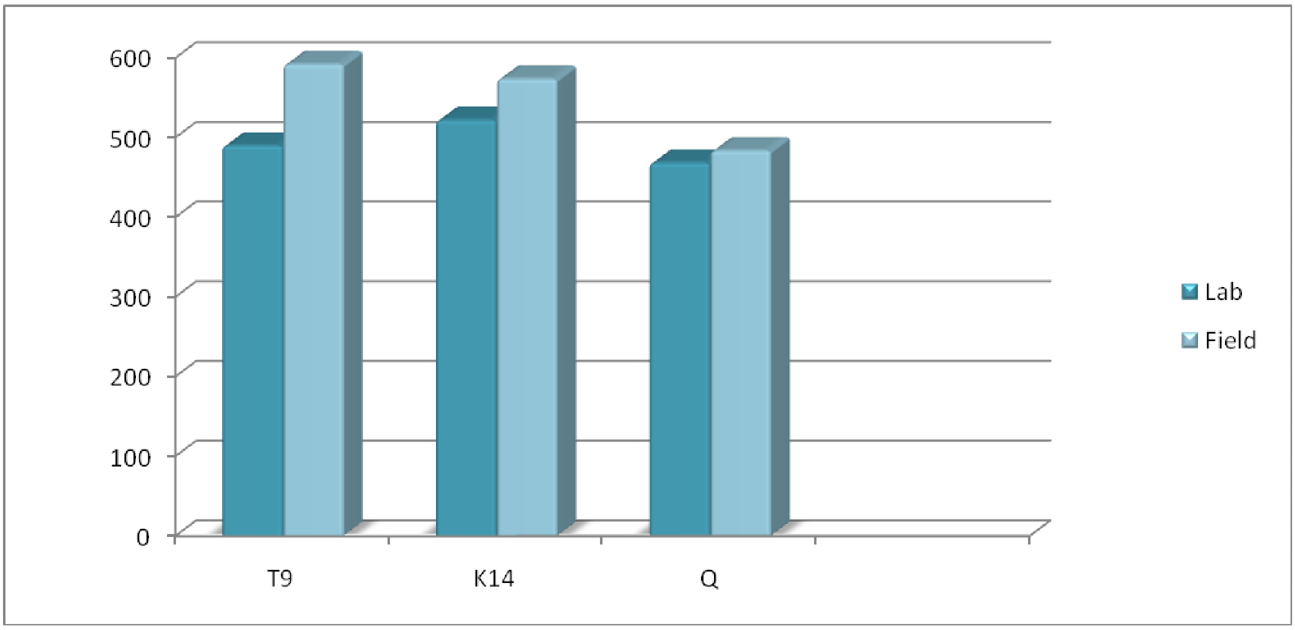
What ~~is~~ system do native and non native English speakers ~~users~~ use - T9 or multi-tap?

native-multi-tap	**Non-native-multi-tap**	**native-T9**	**non-native-T9**
0	4	6	0

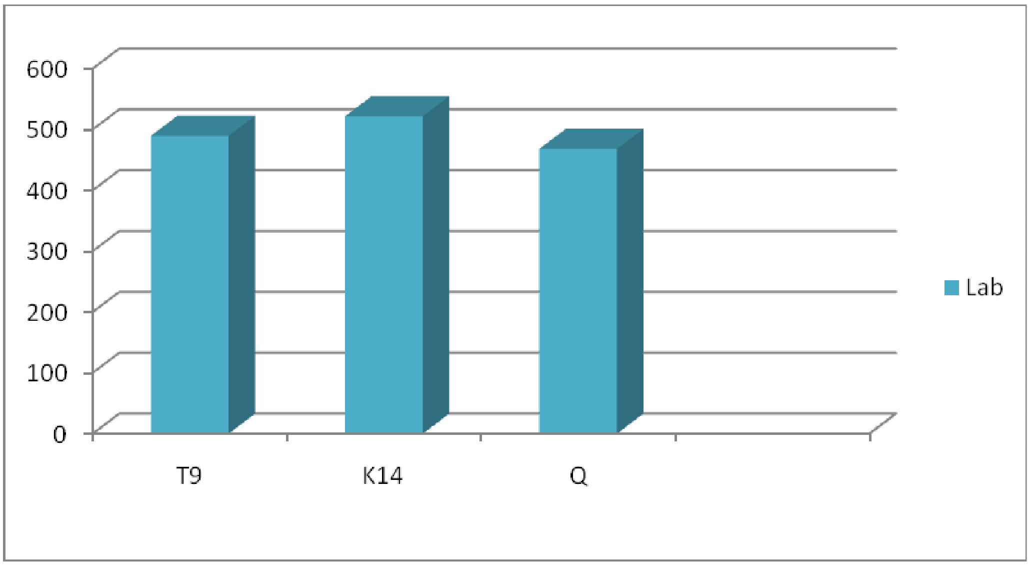
One important factor that emerged was that 100% of non native English speakers used multi tap text input. And all of the participants were using 9 key phones. The reason for this was that they were all used to it and also that 9 key phones were easy to carry as compared to 14 key or QWERTY phones.

3.11 Comparison of Field Vs Lab Experiments

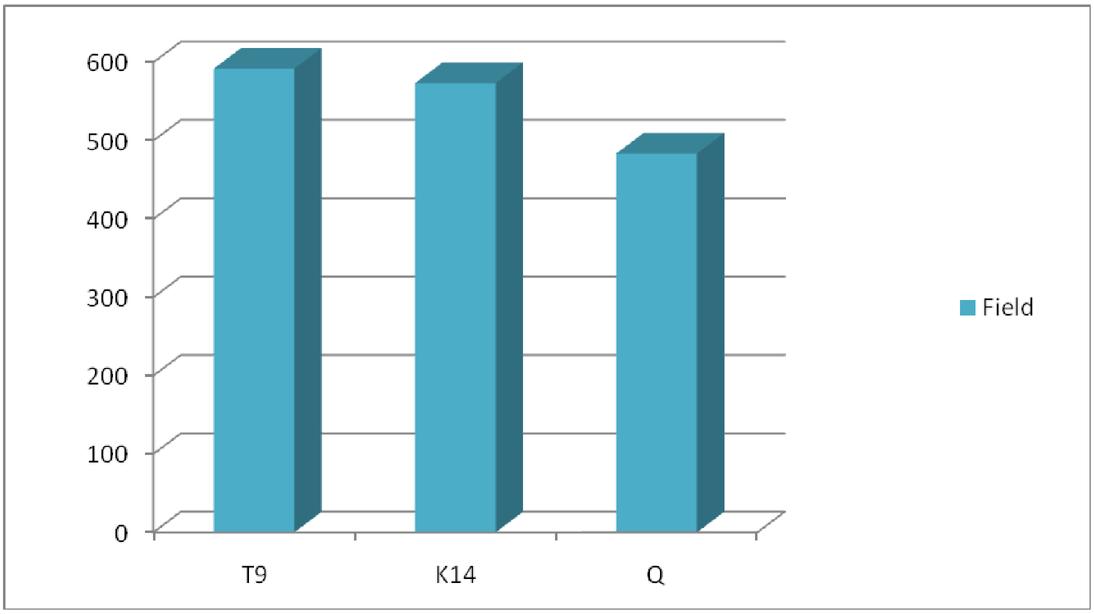
Looking at the graph (a) you will find that users took more time to complete the task in the field than in the lab. Hence it proves the point that there are differences in the results when you conduct two similar experiments in the lab and in the field. Field testing is more close to reality. As in this case users were walking on a path where they had to be aware of their surroundings and also there was noise which contributed to the fact that they took more time to complete the task as compared to lab setting.



Graph a:- Average time taken in seconds to enter 10 phrases in lab vs. Field

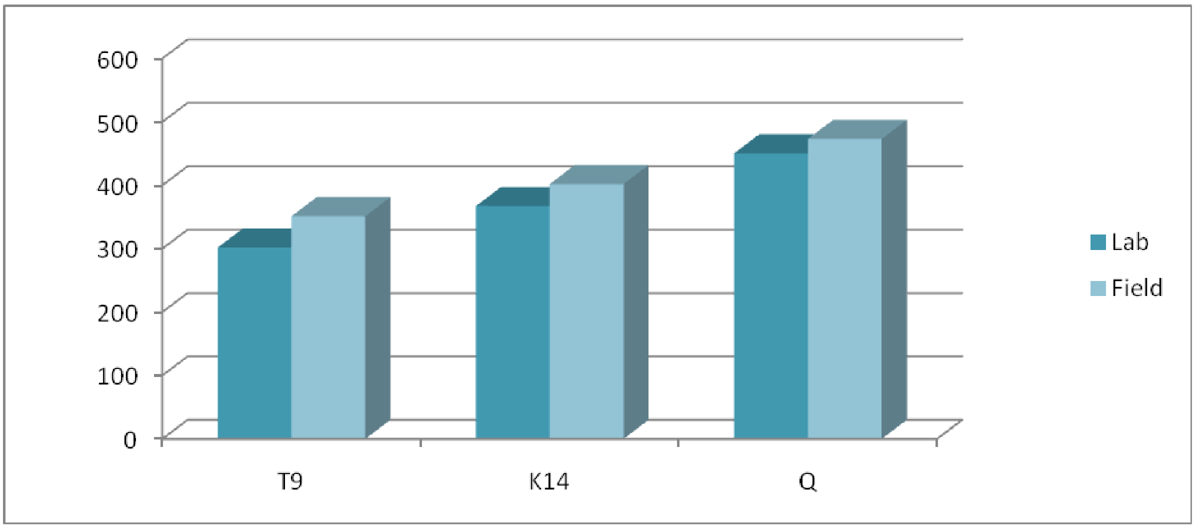


Graph b:- Average time taken in seconds to enter 10 phrases in lab

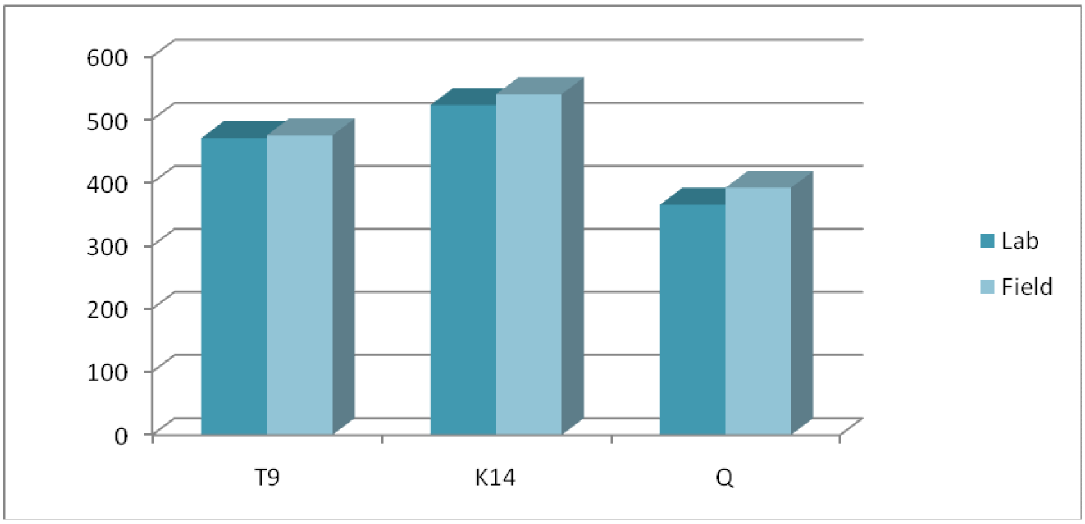


Graph c:- Average time taken in seconds to enter 10 phrases field

By comparing performance of users in lab only as shown in graph B you will find that QWERTY key pad was more efficient than K14 and 12-key phone, and K14 was least efficient among the three. But when you move to field experiments, QWERTY still remained the most efficient of them all but instead of K14, the 12-key phone was least efficient of all this can seen in Graph c.



Graph d:- Average time taken in seconds to enter 10 phrases in lab &. Field By native English Speakers



Graph e: - Average time taken in seconds to enter 10 phrases in lab &. Field by non native English speakers

Since 50% of participants in this experiment were native English speakers and 50 % non-native English speakers. It was thought to be interesting if data from both groups were compared to see if there were any differences in performance. By looking at Graph d and e it was noted that both groups were slower in the field as compared to the lab but their performance was on different key pads. For native English speakers performance deteriorated as number of keys on key pad was increased. But in case of non-native English speakers it was varied - they performed well with QWERTY key pad as compared to 12 key or 20 key.

3.12 Feedback

All participants were asked to complete feedback forms. Feedback form used for this study was a modified version of NASA TLX form.

During study of the feedback and registration forms the following things were observed

- Native English speakers used predictive text input system while non natives used multi tap
- Non native English speakers performance deteriorates as we reduce the number of keys on key pad
- Everyone said they can get used to any keyboard but with practice.
- Based on the stress levels gathered by NASA TLX form, physical and mental stress level was high in the field as compared to the lab.
- Native English speakers physical and mental stress level was low in case of 9 key phones (probably reflecting their experience)
- Non-native English speakers physical and mental stress level was low in case of QWERTY.
- Among three key pads, 9key was most popular because people were use to it.
- People sometimes do compromise performance over shape.
- After the test when asked which phone they will prefer 80% said 9 key 10% said 14 key and 10% QWERTY.

3.13 Conclusions

This chapter has reported a study into the comparison of field and laboratory Usability Testing for Mobile Text Entry. 10 users for lab and 10 for field tests entered set of 10 phrases on three different mobile. Important thing was that order of three set of phrases were changed so that every user can enter different set of phrases on different key pads, in lab and in field. All phrases had equal number of characters. Results were stored on memory card and analysed later. An analysis of lab and field evaluations showed that users were slow in the field as compared to the lab. One important factor for this may be the fact that there is an extra mental load while performing task in field but still it shows that it is important to conduct study in field.

Other observations were the difference in performance of native and non-native English users. Non native English speakers were slow with 14 key keypads as compared to other two keypads this could be because English was not their first language and that a fewer number of keys was creating more confusion. This hypothesis was supported by their better performance with QWERTY key pad as it was less confusing for them.

This difference in performance may be linked with difference in language and since different language means different culture so on the basis of observations in this pilot study it was decided to do further reading on cultures and how they affect usability This was felt necessary and the correct way forward to prove its effects on usability and then try to repeat this study and compare difference in performance between native and non native people along with field versus lab usability evaluations and prove that these results are statistically significant.

Important Note

Unfortunately data gathered during pilot study was lost before full analysis could be done because of an unfortunate incident but data security was kept in mind from that point onwards.

Chapter 4

Study into Cultural Bias

Mobile phones have gained popularity in developing world in the past few years. As mobile markets in developed countries have become almost saturated there are still lots of opportunities in developing world countries like China and India. They have huge populations not having access to mobile phones, but things are changing at a fast rate if you look at the statistics in India alone there are approximately 69 million GSM users and other 29.28 Million CDMA users. Keeping in mind that the total population of India is estimated to be 1 billion there is still a lot of scope for future growth, and this is where most of the companies will focus in future.

But moving to developing countries poses a challenge for the mobile industry because now they will be dealing with users whose first language is not English and who are not literate. It becomes more important to understand language and its effects from usability point of view. It is easy to understand the language if we have an idea about the culture associated with that language. This study will explain what a culture is and how it is important. It will then focus on language effects on usability and then conduct an experiment to prove that difference in languages do have effect on usability results.

Now researchers are seeking ways to understand the target users and the best way to do it is by having ~~a~~ firsthand experience of target users [41]. This research shows that it is important to understand target users for effective research and the best way to do it is by understanding their culture.

4.1 Culture

Culture generally refers to patterns of human activity and the symbolic structures that give such activities significance and importance [42]. Cultures can be understood as systems of symbols and meanings that even their creators contest, that lack fixed boundaries, that are constantly in flux, and that interact and compete with one another [43].

Culture can be defined as all the ways of life including arts, beliefs and institutions of a population those are passed down from generation to generation. Culture has been called "the way of life for an entire society [44]. As such, it includes codes of manners, dress, languages, religion, rituals, games, norms of behaviour such as law and morality, and systems of belief as well as the art.

From an HCI point of view Ford [45] defines culture as ‘the patterns of thinking, feeling, and acting that influence the way in which people communicate among themselves and with computers’. This definition is also applicable to mobile interactions.

4.2 Culture and Mobile Phones

It is a proven fact that mobile device plays different roles in different cultures. For example in one of the studies done in Japan and another in Korea proved that in Korea people are using internet on mobile phones primarily for downloading games, music and other applications but in Japan it was mostly used for sending and receiving e-mails.[46]

If we look at the use of mobile phone in India or Pakistan we will find that mobile phones are being used in totally different way. For example, in many villages where only a few people can afford mobile phones they share it with others, people can call them to get in touch with relatives in that village by agreeing a time to call or people in that village can call someone by paying the exact amount of outgoing calls. In some parts, people are using them as public call facilities where anyone can use it by paying a fixed amount per minute.

Other unique way of using a mobile phone to communicate in rural areas where people are short of cash is missed calls. For example, one missed call may mean that the person has reached destination safely. Two missed calls mean that he is on the way. This way of communication is very popular in rural areas of India and Pakistan. SMS is also widely used because it is very cheap. But the biggest hurdle in many cases is that majority of population do not understand English or in extreme cases cannot read or write any language [47]. In Japan, small sizes of phone and decorations are important. In China, game playing is important and in USA the functionally of the phone is important [48].

This shows that there are differences in cultures when it comes to using mobile devices. Now if we consider definition of culture mentioned above we will find that these differences in uses can be because of the following reasons

- Codes of Manners
- Language
- Religion
- Rituals
- Games
- Norms of behaviour
- Systems of belief

In terms of cultural affecting the use of a product, one interesting example was mentioned from Honlod [49] while investigating use of a German washing machine in India. He identifies eight cultural factors that must be considered in any investigation of the context in which the product issued:

- objectives of the users
- characteristics of the users
- environment
- infrastructure
- division of labour
- organization of work
- mental modes based on previous experience and
- tools

Neilson also recommended travelling to country for conducting usability research related to that culture and also employing local staff [50].It is also statistically proven that the researcher conducting research tends to find more usability issues if they are working with people with same cultures. [51]

In developing countries, language and literacy are barriers that prevent many people from using simple applications like a phonebook on mobile phones. The traditional alphabetical organization is not good enough for low-literate users who either do not know or have forgotten the alphabetical order of any script [52] so some researchers have proposed special interaction systems for developing countries. In the coming years cultural effects on usability testing will become significant.

4.3 Text Input and Cultures

By looking at the success of text messaging and how popular it is worldwide this fact pushes us to think further of contextual and cultural factors on IT products. Take the example of china it is very difficult for users to enter text using a key pad as compared to people in UK because of language but it is still gaining popularity.

One important thing that came to light in pilot study was that almost all of the participants who had English as their first language were using predictive text entry system and all participants whose first language was not English were using multi tap text entry system. The reason for non-native English speakers using multi tap was that they had to send text in their mother language and it was not easy while using predictive text entry system. This shows that there were some differences in behaviour when it comes to entering text in mobile phones. Although better predictive or text entry techniques improves performance [53], it does not guarantee that it will be used fully for people entering text in Chinese or Hindi [54]. Researchers have pointed out that due to the complicated input process for Asian languages, it is more challenging to develop a local version of text entry system.[55]

Although it is possible to send text messages in scripts such as Arabic, Chinese, Japanese, and Thai, the Latin alphabet is widely used by testers around the world [56] - this minimises the effort required to enter text. Most of the Pakistanis use Latin alphabets to send messages in their own language.

4.4 Culture and Usability

Giving culture a priority is very important especially when designing for areas where people have to walk long distances just to charge their mobile phones. Giving language and literacy factors importance in usability evaluations will make products more popular in developing countries and this will bring those people more close to modern technology.
Since another important part of this study was to compare lab and field usability evaluations by carefully reviewing what we have discussed so far, we will clarify the fact that field studies become more important in this context as they will be more useful in understanding different cultures and their effect on usability. This fact has also been recommended by others [57].
Cultures have a well known effect on usability so in order to investigate it further the fact that people in different cultures perform differently and findings of one culture cannot be applied to another it was decided to conduct an experiment. It was decided to select people from two different groups, in this case native and non native English speakers. Although non native English speakers speak and understand English, their level of understanding was different in each case and each group performance level will be different if they are made to do similar tasks. Aim was to investigate that there is a significant difference in performance.

4.5 Experimental Plan

This experiment was designed on similar lines to the pilot study experiment but this time more care was taken in collecting data and storing it. Before the start of experiment, the hypothesis was that users will be slower in the field as compared to the lab and there will be a significant difference in performance of native and non-native English Speakers.

4.6 Location of Experiment

Experiments were conducted at two places. First in an office building which has similar environment as a lab with no distraction and controlled variables. Second experiment was conducted in a garden with people walking and lots of distractions around.

4.7 Number of participants

There were a total of 8 participants in experiment. They were selected on following categories

- Native English speakers 4
- Non Native English Speakers 4

Out of these 8, 6 were male and two females. Average age of all participants was 33.2 years

4.8 Equipment Used

For this experiment mobile devices selected were

- Nokia N70

Two groups of phrases containing equal number of characters were chosen. Each group contained ten phrases which are as follows:

Set A:-

- I agree with you
- play it again Sam
- the living is easy
- the power of denial
- have a good weekend
- all work and no play
- I like to play tennis
- hair gel is very greasy
- my bike has a flat tire
- be home before midnight

Set B:-

- fish are jumping
- rain rain go away
- the cotton is high
- do not say anything
- life is but a dream
- buckle up for safety
- it looks like a shack
- he is still on our team
- do not walk too quickly
- I will meet you at noon

It was ensured that every user enters a different set of phrase for different keypads.
All phrases were arranged in ascending order based on the number of characters in each phrase. Total numbers of characters in each set were equal. Phrases were read out to participants during experiment. The order in which each user was supposed to enter Phrases is as follow

User Number	Phrases to Enter on 9 Key	Phrases to Enter on QWERTY
1	A	B
2	B	A
3	A	B
4	B	A
5	A	B
6	B	A
7	A	B
8	B	A

4.9 Text Input System for Experiments

Text input system used for this experiment was the same as used in pilot study.

4.10 Recording of Results

A J2ME application was installed on the N70 and Palm Tungsten C Phones to record results. It was designed to store data on memory card of the phones in text format and started recording the time taken to type phrase, starting by first key pressed and then logging time every time a space key was pressed it also recorded any errors made by the participants.
Every participant was suppose to correct any errors made during the experiment , press space key at the end of every sentence and clear the screen after the end of sentence.

4.11 Pre Experiment Briefing

All participants were briefed about experiments and were given ample time to practice before actual start of experiment. For the field experiment all participants were given tour of the area so that they were all familiar with the area. It was made sure that all the participants understood experiment and started only once they were ready. They were given following instructions.

4.12 Experiment plan

1) Each user will start the experiment by entering a user ID and then press options button and select 12-key phone/QWERTY option.

2) A blank screen will appear then users will be told to enter text when they are ready.

3) They will finish each sentence by pressing space key. It is compulsory.

4) After that they will press options button and select clear

5) They will repeat step 5-6 for every sentence

4.13 Lab Experiment

Lab experiment was conducted in a room where they had no disturbance. Setup of the lab was similar to the one done in pilot study. Users were allowed to enter text while sitting on the chair. All users were observed from a close distance but it was made sure that observer presence did not hinder them or distract them. Observer was there to ensure that steps were being followed so that data gathered in authentic manner and does not result in wrong interpretation. If a mistake was notice participants were requested to correct that. It was made sure that participants do not feel under pressure. At the end of each session data was transferred to computer and saved on multiple locations so that it remains safe for future use.

Lab Settings

4.14 Field experiment

Like last experiment this experiment was conducted in a park. Participants had to move on a path while performing tasks which diverted their attention. They had to vary speed, change directions. Route was such that they had to be aware of their surroundings while completing the task. It was a good way to depict a real world situation. Only difference between this location and the one used in pilot study was that in pilot study field location was busier than present location. There was no construction work going on no traffic passing close by and very less human traffic as well in the park. As it was not possible to travel to Glasgow and conduct test it was decided to go ahead with this new location but time selected as such that there are lots of people, and dogs around so that maximum distractions can be replicated but still it was not similar to pilot study field location and this can be noticed in the final results.

Field Experiment Location

Path used by participants

4.15 Numerical Analysis

Data gathered from the experiments was compared as follows:

- Data from 12-key phone indoor was compared with QWERTY indoor.
- Data from 12-key phone outdoor was compared with QWERTY outdoor.
- Data from 12-key phone Native indoor was compared with 12-key phone Non-Native indoor.
- Data from 12-key phone Non-Native outdoor was compared with 12-key phone Non-Native outdoor.
- Data from QWERTY Native indoor was compared with QWERTY Non-Native indoor.
- Data from QWERTY Non-Native outdoor was compared with QWERTY Non-Native outdoor.

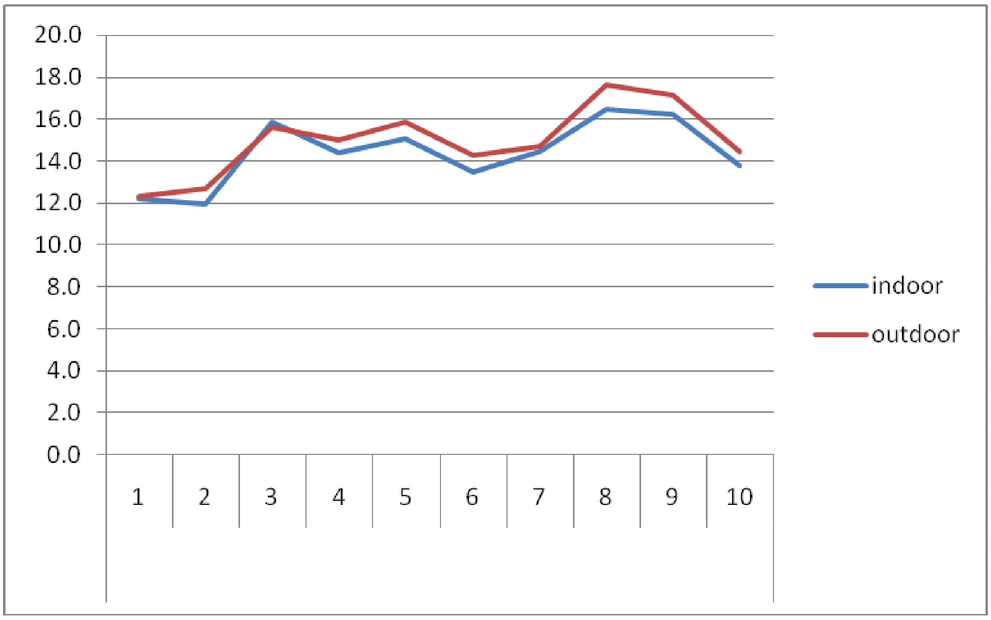
In order to prove that the data gathered was statistically significant Anova with pair wise Tukey-HSD analysis test was performed which is a method specially used for multiple comparisons [52]

12-key phone VS QWERTY

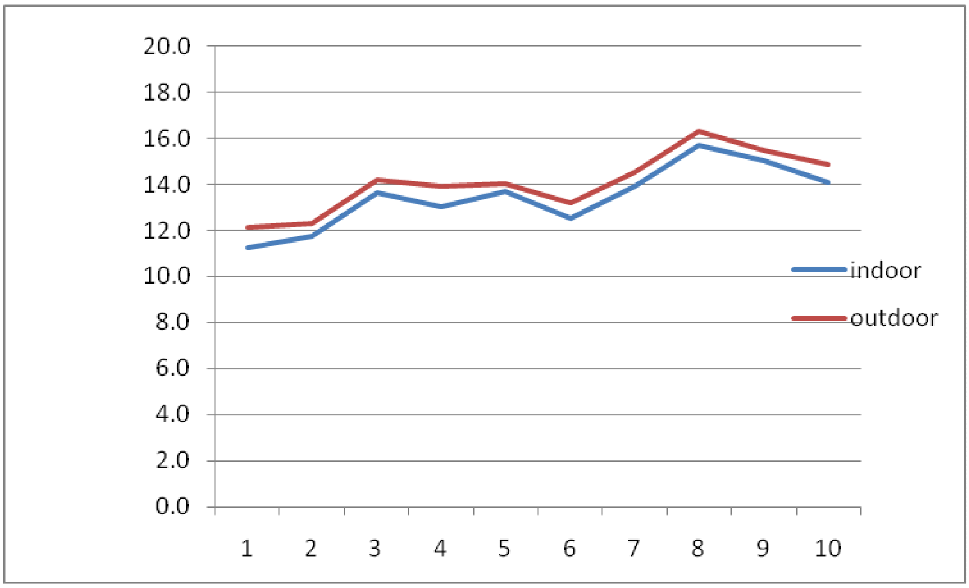
First comparison was done between 12-key phone indoor, 12-key phones outdoor,
in terms of difference in performance of users using 12-key phone, the mean of 12-key phone indoor was 14.38; the mean of 12-key phone outdoor was 14.96. A Tukey two tailed pairs test gives p=0.0001 with n=80 pairs, showing a clearly significant result.

Second comparison was done between QWERTY phone indoor, QWERTY phone out door. In terms of difference in performance of users using QWERTY phone, the mean of QWERTY phone indoor was 13.46, the mean of QWERTY phone outdoor was 14.07. A Tukey two tailed pairs test gives p=0.0001 with n=80 pairs, showing a clearly significant result.

This shows that users were slower in field as compared to lab. Following are the graphs obtained based on comparisons between 12-key phones and QWERTY :



A: Graph for Indoor Vs Outdoor Experiment using 12-key phone



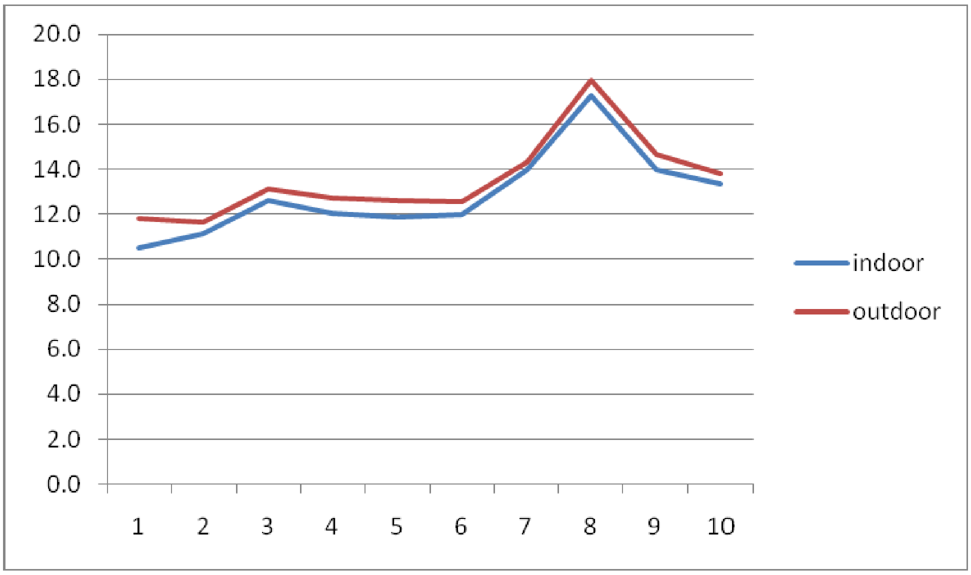
B: Graph for Indoor Vs Outdoor Experiment using QWERTY

4.16 12-key phone Native VS Non Native

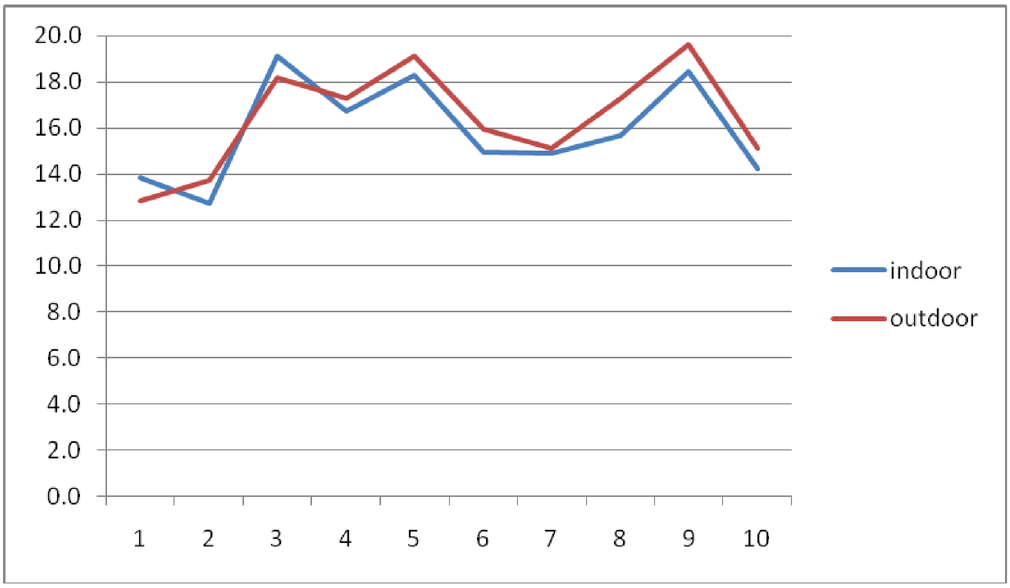
First comparison was done between 12-key phone Native indoor, 12-key phone non native indoor, In terms of difference in performance of native and non native users using 12-key phone indoor, the mean of native 12-key phone indoor was 12.91, the mean of 12-key phone non native indoor was 15.89. A Tukey two tailed pairs test gives p=0.0001 with n=40 pairs, showing a clearly significant result.

Second comparison was done between 12-key phone Native outdoor, 12-key phone non native outdoor. In terms of difference in performance of native and non native users using 12-key phone out door, the mean of native 12-key phone outdoor was 13.56, the mean of 12-key phone non native outdoor was 16.41. A Tukey two tailed pairs test gives p=0.0001 with n=40 pairs, showing a clearly significant result.

Following are the graphs obtained based on comparisons between 12-key phone native indoor and outdoor, 12-key phone non-native indoor and outdoor



C: Graph for Indoor Vs Outdoor Experiment using 12-key phone for Native English Speakers



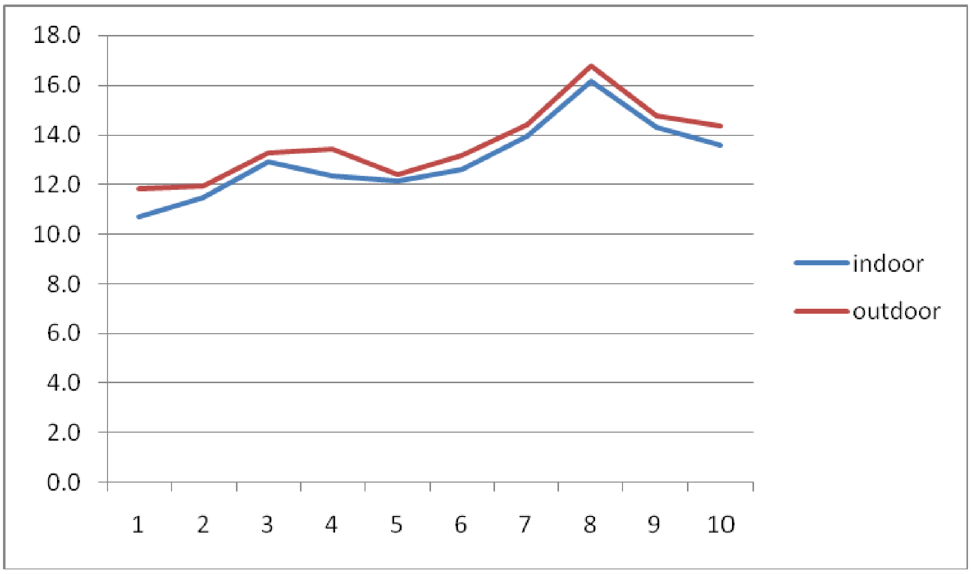
D: Graph for Indoor Vs Outdoor Experiment using 12-key phone for Non-Native English Speakers

4.17 QWERTY Native VS Non Native

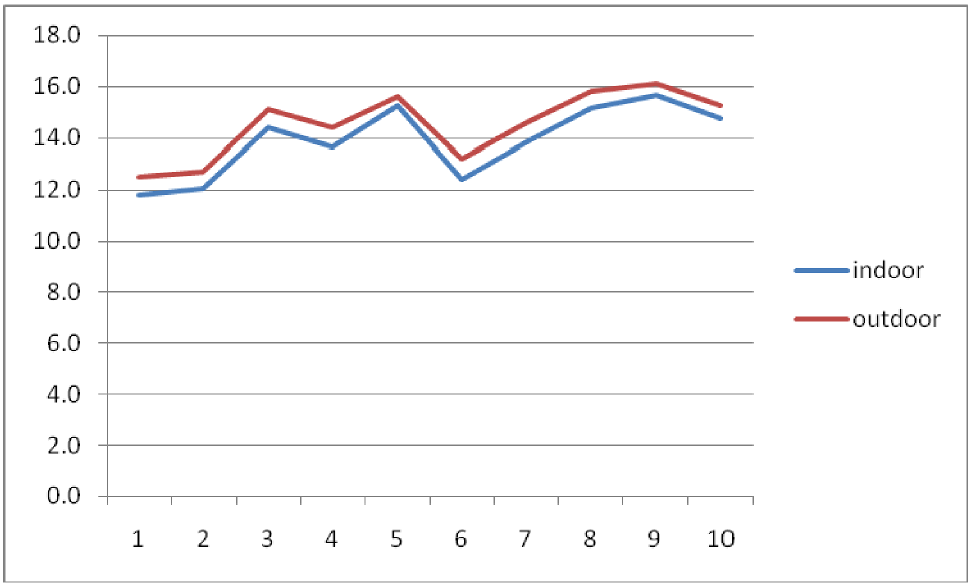
First comparison was done between QWERTY phone Native indoor, QWERTY phone non native indoor. In terms of difference in performance of native and non native users using QWERTY phone indoor, the mean of native QWERTY phone indoor was 13.05, the mean of QWERTY phone non native indoor was 13.895. A Tukey two tailed pairs test gives p= 0.026885 with n=40 pairs, showing a non significant result.

Second comparison was done between QWERTY phone Native outdoor, 12-key phone non native outdoor. In terms of difference in performance of native and non native users using QWERTY phone out door, the mean of native QWERTY phone outdoor was 13.64, the mean of QWERTY phone non native outdoor was 14.51. A Tukey two tailed pairs test gives p= 0.022830 with n=40 pairs, showing a non significant result.

Following are the graphs obtained based on comparisons between QWERTY phone native and QWERTY phone non-native.



E: Graph for Indoor Vs Outdoor Experiment using QWERTY for Native English Speakers



F: Graph for Indoor Vs Outdoor Experiment using QWERTY for Non-Native English Speakers

4.18 Observations

During this study it was observed that there is a significant difference in performance of participants in lab and in field. Users were much quicker in lab setting as compared to field. It is very obvious keeping in mind that in the field they had to be aware of their surroundings and had to perform their task while walking in park.

Native English speakers were quicker in completing their tasks as compared to non-native English speakers. During this study it was observed that for non native English speakers it was hard to remember the sentence and sometimes had to spell the word for them so that that they can type it correctly. They were allowed to ask correct spellings of the word if they were confused so that they can type correctly but this resulted in overall slow performance when compared with native English speakers.

One more observation about non-native English speakers was that they were able to type quicker using QWERTY as it was less confusing compared to 12 key phone keypad. One other factor affecting their performance, on 12 key phone keypad was the fact that almost all of them were using multi tap instead of predictive system and while typing in slow they had to spend some time figuring out correct key. Overall performance of non native English speakers improved when they moved from 12key pad phone to QWERTY key pad phone.

4.19 Results

This chapter has reported a study into comparison of native and non-native English speakers performance entering text on 12 key and QWERTY mobile phone key pads. 8 users took part in Lab experiment, 4 were native English speakers and 4 were non native English speakers. Similarly 8 users took part in field experiment, 4 were native English speakers and 4 were non native English speakers.

After doing analysis of the results obtained during this study following points were statistically proven.

- Users were slower in field as compared to lab in using both 9 key and QWERTY Key Pads.
- Users took less time in completing task using QWERTY than T9 both in lab and in Field
- Non native English speakers were slower in completing task using T9 as compared to native
- Non native English speakers were slower in completing task using QWERTY Key pad as compared to Native.

In this study, results were not as clear as they were in the first study in terms of lab and field comparisons. After looking at the field experiment locations for both settings it was observed that in second study park chosen was quieter as compared to first study and did not have stairs . It also shows that these things do make an impact on usability results and support our hypothesis.

The difference in performance noted during this study can better be understood in terms of cognitive load theory which is the level of effort associated with thinking and reasoning. By looking at the experiment we will find that cognitive load for non-native English speakers was high as compared to native English speakers because non native English speakers had to do more effort as all instructions were in English. Same was the case between 12 key and QWERTY Key pad, cognitive load was high for 12 key as users were required to find keys with appropriate characters for the experiment while it was much simpler in case of QWERTY. Similarly cognitive load for users performing tasks in the field was much higher as compared to lab and this fact is proven by the data gathered during experiment. This experiment also proved that cognitive load also effects performance.

5. Conclusion and Future Work

This study was initially aimed at finding the differences on the text entering performance of users when number of keys on key pads are changed and to get an idea about what are the differences when we conduct same study but at different locations. In that study it was noticed that number of keys on key pads did had effect on the overall performance of users but most interesting observations was when a comparison was made between usability testing conducted in the field and in the lab and also comparison between native and non native English speakers.

While conducting usability testing in the field it was observed that users were distracted by passing by traffic, people and construction work taking place on one corner of garden. This made it more challenging for them to complete the task and naturally as was shown that they took more time to complete the task. So it was decided to repeat the study and see if difference in performance was significant.

In terms of native and non native English speakers observations was made during the experiment that performance of both groups were different on different key pads. This became clearer when data was separated on the basis of native and non native English speakers. This was also part of second study so that it can be proved that there is a significant difference in both cases.

After completion of second phase, Tukey-HSD analysis test was conducted to prove significance of results. Based on analysis, ~~test~~ following observations were made.

There is a significant difference in performance of users in lab and in field which proved our first hypothesis that users are slower performance wise in field as compared to lab. It was also observed that there is a significant difference in performance of native and non native English speakers but number of key pads did not have a significant effect on the performance of native and non native users.

Difference in performance of native and non-native English speakers indicated that cultures do effect ~~on~~ usability evaluations, for example, use of one technology can change if we move between culture to cultures and also the level of expertise and understanding are different specially if some product is designed by people living in a different culture and speaking a different language. It is important to keep in mind that we are developing a product that will be used globally and most of the people in poor countries will not be able to understand it correctly which will be a contributory factor towards the deterioration of their performance . Future studies will focus on the cultural aspects of usability evaluation techniques in more details.

5.1 Future Work

It is a proven fact that in today's increasingly global market, software products must be equally usable across different countries and cultures. This means that product success depends strongly on the international usability of products. Mobile phones are a global success and it is important to have usability techniques that are focused on cross cultural use of mobile devices.

Taking it further a more detailed study on the cultural effects on usability of mobile devices will be conducted. It will be done by performing usability testing in three different cultures and establishing a link between cultures and usability. This study will also elaborate on the importance of conducting usability testing in the target culture itself instead of doing it remotely or getting users from target culture but living in a different culture. This study will help us build general guidelines for usability testing with cultural effects as a focal point. This will increase satisfaction level and success rate of products and will make new products easy to use by a wider range of people all across the world.

Another important part of future work is field usability since a lot has been argued for and against it but both usability techniques have their own plus points. Choosing between the two may depend on the usability testing criteria and environment a product will be used. Although people have talked about harvesting the good of both techniques it is important from mobile usability point of view that there is a system that can bring the best of both techniques under one roof. One study will focus on identifying factors that affect performance of users in field and try to regenerate same environment in lab. This has not only given researches more control but also make usability more close to real world and data obtained will help researchers make suggestions that will improve overall user experience.

Experiment Data

12 Key Vs QWERTY

Sample 1	Sample 2	Sample 3	Sample 4
12.44	12.3	11.4	11.8
11.3	11.4	10.3	11.2
23.7	18.5	13.7	14.3
17.8	19.2	15.3	15.7
23.7	24.3	18.4	18.1
12.8	13.9	10.8	12.1
13.7	13.6	12.7	13.1
14.2	14.8	13.8	14.5
19.1	20.3	14.5	14.7
12.06	13.7	13.6	14.9
10.44	11.2	10.2	11.1
10.15	12.7	9.8	10.2
13.9	14.3	12.7	13.4
12.2	12.3	11.4	11.8
12.4	12.4	12.3	12.4
12.8	13.7	12.7	13.3
13.8	13.9	14.5	15.6
18.5	19.2	18.9	18.4
14.3	15.4	14.4	15.1
14.55	16.2	15.3	15.9
17.5	12.7	11.1	11.8
14.7	15.1	13.9	14.4
18.4	18.6	15.7	16.8
15.12	15.7	14.2	15.4
16.35	16.8	15.4	16.7
16.6	16.8	13.1	13.5
13.3	13.8	12.4	12.9
9.9	13.8	10.1	11.7
17.1	18.2	16.4	17
16.68	16.68	15.9	16.3
14.9	15.2	14.5	15.2
14.8	15.8	14.2	14.9
20.4	21.2	15.4	15.8
21.7	21.8	13.7	14.7
20.7	22.8	14.8	15.2
17.7	19.5	13.1	13.9
18.8	19.1	15.8	16.7
20.1	21.3	17.9	18.5
23.2	24.5	17.3	17.6
13.7	13.9	14.2	13.8
10.3	12.4	10.4	12
11.1	11.7	11.3	11.6
12.9	13.5	13.5	14.1
11.8	11.9	12.1	12.8
12.2	12.8	12.7	12.5

11.8 14.1 17.5 14.2 13.4 9.8 11.8 12.1 12.3 12.8 11.7 13.9 16 13.9 13.8 11.1 10.25 12.3 11.8 11.2 12.45 14.2 17.3 13.5 13.8 10.9 11.4 13.2 12.35 11.3 12.1 13.7 18.2 14.3 12.1	12.1 14.35 18.2 14.6 13.7 11.7 12.3 12.8 12.5 13.8 12.4 14.3 16.2 14.7 14.6 11.9 11.2 12.7 12.7 11.5 12.8 14.5 17.8 13.7 14.5 11.3 11.3 13.4 13.8 12.4 12.9 14 19.7 15.7 12.2	12.2 13.9 17.8 14.7 12.7 10.1 12.1 12.3 12.7 13.2 12.4 14.1 15.4 14.2 14.1 11.3 10.7 12.4 12.1 11.3 12.7 14.5 16.3 14.5 14.1 11.1 11.8 13.5 12.6 11.4 13.2 13.4 15.2 13.9 12.8	12.36 14.7 18.1 14.9 13.8 11.7 12.5 13.4 14.7 13.7 12.9 14.6 16.1 15.2 14.7 11.7 11.3 12.6 12.8 11.6 13.2 14.9 16.7 13.7 15.6 11.9 12.3 12.9 13.3 11.8 14.1 13.5 16.2 15.3 13.8

Data Summary

	1	2	3	4	Total
N	80	80	80	80	320
ΣX	1150.34	1197.13	1076.5	1125.96	4549.93
Mean	14.3793	14.9641	13.4563	14.0745	14.2185

Σx^2	17406.0126	18726.1549	14801.15	16138.7496	67072.0671
Variance	10.9492	10.2804	3.9936	3.6889	7.4569
Std.Dev.	3.309	3.2063	1.9984	1.9207	2.7307
Std.Err.	0.37	0.3585	0.2234	0.2147	0.1527

ANOVA Summary

Source	SS	df	MS	F	P
Treatment [between groups]	94.6847	3	31.5616	19.2	<.0001
Error	389.6639	237	1.6442		
Ss/Bl	1894.3966	79			
Total	2378.7452	319			

Ss/Bl = Subjects or Blocks depending on the design.
Applicable only to correlated-samples ANOVA.

Tukey HSD Test

HSD[.05]=0.53; HSD[.01]=0.64 M1 vs M2 P<.05 M1 vs M3 P<.01 M1 vs M4 nonsignificant M2 vs M3 P<.01 M2 vs M4 P<.01 M3 vs M4 P<.05	M1 = mean of Sample 1 M2 = mean of Sample 2 and so forth. HSD = the absolute [unsigned] difference between any two sample means required for significance at the designated level. HSD[.05] for the .05 level; HSD[.01] for the .01 level.

Native Vs Non native indoor out door

Sample 1	Sample 2	Sample 3	Sample 4
12.44	12.3	10.3	12.4
11.3	11.4	11.1	11.7
23.7	18.5	12.9	13.5
17.8	19.2	11.8	11.9
23.7	24.3	12.2	12.8
12.8	13.9	11.8	12.1
13.7	13.6	14.1	14.35
14.2	14.8	17.5	18.2
19.1	20.3	14.2	14.6
12.06	13.7	13.4	13.7
10.44	11.2	9.8	11.7
10.15	12.7	11.8	12.3
13.9	14.3	12.1	12.8
12.2	12.3	12.3	12.5
12.4	12.4	12.8	13.8

12.8	13.7	11.7	12.4
13.8	13.9	13.9	14.3
18.5	19.2	16	16.2
14.3	15.4	13.9	14.7
14.55	16.2	13.8	14.6
17.5	12.7	11.1	11.9
14.7	15.1	10.25	11.2
18.4	18.6	12.3	12.7
15.12	15.7	11.8	12.7
16.35	16.8	11.2	11.5
16.6	16.8	12.45	12.8
13.3	13.8	14.2	14.5
9.9	13.8	17.3	17.8
17.1	18.2	13.5	13.7
16.68	16.68	13.8	14.5
14.9	15.2	10.9	11.3
14.8	15.8	11.4	11.3
20.4	21.2	13.2	13.4
21.7	21.8	12.35	13.8
20.7	22.8	11.3	12.4
17.7	19.5	12.1	12.9
18.8	19.1	13.7	14
20.1	21.3	18.2	19.7
23.2	24.5	14.3	15.7
13.7	13.9	12.1	12.2

Data Summary

	1	2	3	4	Total
N	40	40	40	40	160
Σ**X**	635.49	656.58	514.85	540.55	2347.47
Mean	15.8873	16.4145	12.8713	13.5138	14.6717
ΣX^2	10637.8251	11279.0124	6768.1875	7447.1425	36132.1675
Variance	13.8881	12.861	3.6263	3.6483	10.6341
Std.Dev.	3.7267	3.5862	1.9043	1.9101	3.261
Std.Err.	0.5892	0.567	0.3011	0.302	0.2578

ANOVA Summary

Source	SS	df	MS	F	P
Treatment [between groups]	363.8953	3	121.2984	24.46	<.0001
Error	580.3202	117	4.96		

Ss/Bl	746.6058	39
Total	1690.8212	159

Ss/Bl = Subjects or Blocks depending on the design.
Applicable only to correlated-samples ANOVA.

Tukey HSD Test

HSD[.05]=1.3; HSD[.01]=1.58 M1 vs M2 nonsignificant M1 vs M3 P<.01 M1 vs M4 P<.01 M2 vs M3 P<.01 M2 vs M4 P<.01 M3 vs M4 nonsignificant	M1 = mean of Sample 1 M2 = mean of Sample 2 and so forth. HSD = the absolute [unsigned] difference between any two sample means required for significance at the designated level. HSD[.05] for the .05 level; HSD[.01] for the .01 level.

QWERTY Native Vs Non Native in door and out door

Sample 1	Sample 2	Sample 3	Sample 4
10.4	12	11.4	11.8
11.3	11.6	10.3	11.2
13.5	14.1	13.7	14.3
12.1	12.8	15.3	15.7
12.7	12.5	18.4	18.1
12.2	12.36	10.8	12.1
13.9	14.7	12.7	13.1
17.8	18.1	13.8	14.5
14.7	14.9	14.5	14.7
12.7	13.8	13.6	14.9
10.1	11.7	10.2	11.1
12.1	12.5	9.8	10.2
12.3	13.4	12.7	13.4
12.7	14.7	11.4	11.8
13.2	13.7	12.3	12.4
12.4	12.9	12.7	13.3
14.1	14.6	14.5	15.6
15.4	16.1	18.9	18.4
14.2	15.2	14.4	15.1
14.1	14.7	15.3	15.9

11.3	11.7	11.1	11.8
10.7	11.3	13.9	14.4
12.4	12.6	15.7	16.8
12.1	12.8	14.2	15.4
11.3	11.6	15.4	16.7
12.7	13.2	13.1	13.5
14.5	14.9	12.4	12.9
16.3	16.7	10.1	11.7
14.5	13.7	16.4	17
14.1	15.6	15.9	16.3
11.1	11.9	14.5	15.2
11.8	12.3	14.2	14.9
13.5	12.9	15.4	15.8
12.6	13.3	13.7	14.7
11.4	11.8	14.8	15.2
13.2	14.1	13.1	13.9
13.4	13.5	15.8	16.7
15.2	16.2	17.9	18.5
13.9	15.3	17.3	17.6
12.8	13.8	14.2	13.8

Data Summary

	1	2	3	4	Total
N	40	40	40	40	160
ΣX	520.7	545.56	555.8	580.4	2202.46
Mean	13.0175	13.639	13.895	14.51	13.7654
ΣX^2	6879.13	7540.1696	7922.02	8598.58	30939.8996
Variance	2.5876	2.5456	5.1072	4.5378	3.9133
Std.Dev.	1.6086	1.5955	2.2599	2.1302	1.9782
Std.Err.	0.2543	0.2523	0.3573	0.3368	0.1564

ANOVA Summary

Source	SS	df	MS	F	P
Treatment [between groups]	45.8623	3	15.2874	8.34	<.0001
Error	214.5344	117	1.8336		
Ss/Bl	361.8151	39			
Total	622.2118	159			

Ss/Bl = Subjects or Blocks depending on the design.
Applicable only to correlated-samples ANOVA.

Tukey HSD Test

HSD[.05]=0.79; HSD[.01]=0.96 M1 vs M2 nonsignificant M1 vs M3 P<.05 M1 vs M4 P<.01 M2 vs M3 nonsignificant M2 vs M4 P<.05 M3 vs M4 nonsignificant	M1 = mean of Sample 1 M2 = mean of Sample 2 and so forth. HSD = the absolute [unsigned] difference between any two sample means required for significance at the designated level. HSD[.05] for the .05 level; HSD[.01] for the .01 level.

Revised Form

Mental Demand

How much mental and perceptual activity was required while doing the task in lab?

low ☐ ☐ ☐ ☐ ☐ ☐ ☐ high

How much mental and perceptual activity was required while doing the task in field?

low ☐ ☐ ☐ ☐ ☐ ☐ ☐ high

Physical Demand

How much physical activity was while doing the task in lab?

low ☐ ☐ ☐ ☐ ☐ ☐ ☐ high

How much physical activity was while doing the task in lab?

low ☐ ☐ ☐ ☐ ☐ ☐ ☐ high

Performance

How successful do you think you were in accomplishing the goals of the task set by the experimenter in lab?

perfect ☐ ☐ ☐ ☐ ☐ ☐ ☐ failure

How successful do you think you were in accomplishing the goals of the task set by the experimenter in field?

perfect ☐ ☐ ☐ ☐ ☐ ☐ ☐ failure

Effort

How hard did you have to complete task in lab?

low ☐ ☐ ☐ ☐ ☐ ☐ ☐ high

How hard did you have to complete task in field?

low ☐ ☐ ☐ ☐ ☐ ☐ ☐ high

Frustration Level

How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task in lab?

low ☐ ☐ ☐ ☐ ☐ ☐ ☐ high

How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task in field?

low ☐ ☐ ☐ ☐ ☐ ☐ ☐ high

References:-

[1] Nielsen, J. (1994). *Usability Engineering*, Academic Press Inc, P: 165.

[2] *http://en.wikipedia.org/wiki/SMS*

[3] members.optusnet.com.au/~webindexing/Webbook2Ed/glossary.htm

[4] K. H. Connelly, K.A. Siek, V.L.Favieres, and G. Bennett, *Planes, Pains, and Phosphorane: Usability Studies in Non-Traditional Environments,* In Proceedings of Interact 2005

5] http://en.wikipedia.org/wiki/Usability

[6] J. Nielsen . *Usability Engineering*, Chapter 6 complete

[7] http://en.wikipedia.org/wiki/Usability

[9] P. Isokoski, I. S. MacKenzie . *Text entry on mobile systems: Directions for the future.* Extended Abstracts of the ACM Conference on Human Factors in Computing Systems - CHI *2001*; P: 495

[10] C. L. James & K. M. Reischel. "*Text input for mobile devices: comparing model prediction to actual Performance*", Proceedings of the SIGCHI conference on Human factors in computing systems, 2001; P: 365-371

[11] R. W. Ha, P.H. Ho, and X. (Sherman) Shen. *SIMKEYS: An Efficient Keypad Configuration for Mobile Communications*, Communications Magazine, IEEE 2004

[12] F. E. Sandnes, H. Thorkildssen, A. Arvei, J. O.Buverud. *Fast text-entry on miniature mobile devices*, Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04) - Track 9

[13] Gong, J. & Tarasewich. *Guidelines for Handheld Mobile Device Interface Design*, Proceedings of the 2004 Decision Sciences Institute (DSI) Annual Meeting; P: 3751-3756

[14] M. Dunlop, A. Crossan. *Dictionary based text entry method for mobile phones* , Proceedings of Second Workshop on Human Computer Interaction with Mobile Devices. 1999

[15] M. D. Dunlop and M. M. Masters. *Pickup Usability Dominates*. International Journal of Mobile Human Computer Interaction. 2009; P: 42-59.

[22] F. E. Sandnes, H. Thorkildssen A. Arvei J. O.Buverud. *Fast text-entry on miniature mobile devices*, Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04) - Track 9

[23] J. Gong, & P. Tarasewich, *Guidelines for Handheld Mobile Device Interface Design*, Proceedings of the 2004 Decision Sciences Institute (DSI) Annual Meeting ; P: 3751- 3756
[24] J. Gong, P. Tarasewich. *Testing Predictive Text Entry Methods with Constrained Keypad Designs*, Proceedings of the SIGCHI conference on Human factors in computing systems 2005; P: 211 - 220
[25]H.R. Stevens,S. Skiena. *Dailing for Documents: An Experiment In Information Theory* , Proceedings of the 7th annual ACM symposium on User interface software and technology; P: 147 - 155
[26] I. S.MacKenzie, H. Kober, D. Smith, T. Jones , E. Skepner , *LetterWise: prefix-based disambiguation for mobile text input*, Proceedings of the 14th annual ACM symposium on User interface software and technology; P: 111-120
[27] J. C. Read, *Deceit, Distortion and Decision – Choosing Phrase Sets for Text Entry Research,* Child Computer Interaction Group (2005)
[28] J.Gong , P.Tarasewich, and I. S.MacKenzie, *Improved word list ordering for text entry on ambiguous keyboards*, Proceedings of the Fifth Nordic Conference on Human-Computer Interaction - NordiCHI 2008; P: 152-161
[29] J.Gong, B.Haggerty, and Tarasewich. *An Enhanced Multitap Text Entry Method with Predictive Next-Letter Highlighting*, CHI '05 extended abstracts on Human factors in computing systems; P: 1399-1402.
[30] I. S. MacKenzie, & R. W.Soukoreff. *Text entry for mobile computing: Models and methods, theory and practice*. Human-Computer Interaction 2002; P: 147-198
[31]S.P Yong, H.H Sung, P.Jaehyun, C.Youngseok **,** *Touch key design for target selection on a mobile phone* ,Proceedings of the 10th international conference on Human computer interaction with mobile devices and services ;P: 423-426
[32] R. Hardy, E. Rukzio, *Touch & interact: touch-based interaction of mobile phones with displays*, Proceedings of the 10th international conference on Human computer interaction with mobile devices and services; P: 245-254
[33] I. S.MacKenzie, *Mobile text entry using three keys*, Proceedings of the Second Nordic Conference on Human-Computer Interaction – NordiCHI 2002; P: 27-34.
[34] P.Isokoski & MacKenzie, I. S. *Text entry on mobile systems: Directions for the future*. Conference on Human Factors in Computing Systems– CHI 2001; P: 495-495.
[35] Soukoreff, I. S.MacKenzie. *Metrics for text entry research: an evaluation of msd and kspc, and a new unified error metric* , Proceedings of the SIGCHI conference on Human factors in computing systems; P:113-120.

[36] C.M. Nielsen, M Overgaard, M.B. Pedersen, J. Stage, S. Stenild, *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: changing roles 2006;P:272-280.

[37] J. Kjeldskov, J. Stage. *New techniques for usability evaluation of mobile systems,* Proceedings of the 8th conference on Human-computer interaction with mobile devices and services; P: 181-186.

[38] P. Luff, C. Heath. *Mobility in Collaboration*, Proceedings of the 1998 ACM conference on Computer supported cooperative work; P: 305 - 314

[39] P.Johnson. *Usability and Mobility; Interactions on the move*, Proceedings of the First Workshop on Human-Computer Interaction with Mobile Devices 1998

[40] J.Kjeldskov, Connor Graham. *A Review of Mobile HCI Research Methods*, Mobile HCI 2003

[41] J.Kjeldskov, M.B.Skov, B.S. Als, and R.T. Høegh, *Is it Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field*, Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles; P: 272-280.

[42] A. Kaikkonen , M. Cankar , T. Kallio , A. Kankainen . *Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing*, Journal of *Usability* Studies November 2005; P. 1-13

[43] H.B. Duh, C. B. Tan, V.H. Chen. *Usability Evaluation for Mobile Device: A Comparison of Laboratory and Field Tests* Proceedings of the 8th conference on Human-computer interaction with mobile devices and services 2006; P: 181 - 186

[44] F. Jambon. *Usability evaluation of mobile devices: laboratory, field or real world?*, ADAMOS Seminar Final ADAMOS Seminar Adaptive User Interfaces - Technology and Usage 15 June 2007;P: 6-7.

[45] E.Hoggan, S.A. Brewster, and J.Johnston, *Investigating the Effectiveness of Tactile Feedback for Mobile Touch screens*. Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems; P: 1573-1582.

[46] K.Yatani. *An Evaluation of Stylus-Based Text Entry Methods on Handheld Devices in Stationary and Mobile Settings*, Proceedings of the 9th international conference on Human computer interaction with mobile devices and services 2007; P: 487-494.

[47] M. D. Dunlop ,M.M. Masters. *Investigating five key predictive text entry with combined distance and keystroke modelling*, Personal and Ubiquitous Computing 2008; P: 589 - 598

[48] I. S.MacKenzie and R. W. Soukoreff. *Phrase Sets for Evaluating Text Entry Techniques.* CHI '03 extended abstracts on Human factors in computing systems; P: 754-755.

[49] Y. Lee, J. Kim, I. Lee, H. Kim .*A Cross –Cultural Study on the Value Structure Of Mobile Internet Usage: Comparison Between Korea and Japan*, Journal of Electronic Commerce Research, VOL. 3, NO. 4, 2002; P: 227-239.

[50] A .Beu, P.Honold, & X. Yuan. *How to Build Up an Infrastructure for Intercultural Usability Engineering*. (2000). The International Journal of Human-Computer Interaction, 12 (3&4); 347-358.

[51] N. L. Hoft. *International User Interfaces*, (1996) ; P: 41 - 73

[52] R. Vatrapu and A.Manuel. *Culture and Usability Evaluation: The Effects of Culture in Structured Interviews*. Journal of usability studies 2006; P: 156-170

[53] I. S.MacKenzie, and R.W.Soukoreff, *Phrase sets forevaluating text entry techniques*, CHI '03 extended abstracts on Human factors in computing systems; P: 754-755

[54] A.Light. *Designing for other cultures: putting Hofstede to bed*. OzChi2008 proceedings.

[55] Card, Moran & Newell, *The Psychology of Human Computer Interaction*, 1983; P: 209-242

[56] A. L Kroeber, and C. Kluckhohn. *Culture: A Critical Review of Concepts and Definitions*; P: 15-65

[57] T. Vöhringer-Kuhnt. *The Influence of Culture on Usability*, Dept. of Educational Sciences and Psychology. Freie Universität Berlin. Berlin, Germany; P: 2-3

[58] P. Honold. *An Empirical Study for the Development of a Framework for the Elicitation of Cultural Influence in Product Usage***,** International Journal of Human-Computer Interaction, 12 (3&4) (2000); P: 327--245.

[59] J. Nielsen. *Usability testing of international interfaces, Designing user interfaces for international use*, Elsevier Science Publishers Ltd., Essex, 1990

[60] K.F. Law and E. Shon. *The Current Mobile Phone Market & Interaction Design in India*, Human-Computer Interaction. New Trends; P: 677-686

[61] R. Raju. *Minimizing Cultural Bias in Usability Testing*, IPCC 2008 proceedings
[62] A. Joshi , N. Welankar , Naveen , K. Kanitkar , R. Sheikh , *Rangoli: a visual phonebook for low-literate users*, Proceedings of the 10th international conference on Human computer interaction with mobile devices and services;P: 217-223
[63] H.Sun. *Exploring Cultural Usability: A Localization Study of Mobile Text Messaging Use* . CHI '03 extended abstracts on Human factors in computing systems; P:670-671