THE HONG KONG
POLYTECHNIC UNIVERSITY
香港理工大學
DEPARTMENT OF COMPUTING

USABILITY METRICS FOR APPLICATION ON MOBILE PHONE

A Dissertation

by

Chun Wah JO

A Thesis Submitted in Partial Fulfillment of the
Requirement for the Degree of
Master of Philosophy
Department of Computing

The Hong Kong Polytechnic University
June, 2005

THE HONG KONG
POLYTECHNIC UNIVERSITY
香港理工大學
DEPARTMENT OF COMPUTING

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

Jo Chun Wah (Name of student)

Abstract

This research is to investigate different usability issues of the software application in smartphone because the software development in smartphone platform will become a trend in our industry. Then we proposed some usability metrics for the software application in this environment.

Firstly, we consolidated and standardized most of the currently available usability metrics. Although many usability metrics have been proposed, they often either lack of information or they are based on a variety of measurement scales and formats. Our usability metric list provides a useful reference for software developers who want to measure the usability of their products more consistently and transparently. It classifies the metrics according to characteristics such as usability attributes they measure, whether the measurement can be automated, and in which phase of product development the metric can be applied.

Secondly, some usability problems in the smartphone platforms were identified which are quite different compared with the other platform such as desktop computer, laptop computer and even PDA. Based on the special situation of the smartphone with very small size screen and high mobility but limited interface for input and output, some constraints for software application developers have been identified which may affect the usability of the application.

Some usability metrics specialized for the smartphone environment were proposed. Our research objective is to use metrics to evaluate the usability of the software application in smartphones. We designed a usability experiment for the software application in this platform in order to verify the proposed usability metrics. At the

same time, we proposed some hypotheses which could show the focus of our research.

Sixty subjects were invited to participate in the experiments in order to investigate which usability problems will substantially affect the usage of software applications in smartphones. The experiment had two parts: Usability Test and Questionnaire. Our research had applied both quantitative and qualitative methods in order to capture the result from different aspects.

By using SPSS, the data from user behaviours and questionnaire were analyzed in order to validate the usability metrics and the hypotheses. As all experiments had been captured by digital video camera, the qualitative information, such as their comments and suggestions, could also be analyzed. Our findings can improve the usability of the software application not only in smartphones but also in other mobile devices which have similar operation environment.

Acknowledgements

I would like to thank Dr. Hareton Leung for his professional guidance, wisdom, encouragement, and support, without which this dissertation would not have been possible. I am also thankful to Dr. Raymond Lee for his valuable support and suggestions about the usability experiment.

I sincerely thank my dissertation committee members for their intellectual support, feedback and encouragement.

I would like to thank my fellow graduate classmates, Liu Wan Tung, Mike Fung, Gary Lam, Teddy Chow, Alex Wong, Perry Lam, Edward Chung, Iu Chun for their help and friendship, which made my MPhil study very enjoyable.

I gratefully acknowledge Grace Mak for reviewing and giving comments to my conference paper. Special thanks to Yvette Lui who generously contributed so much of her time to review the draft of this dissertation.

I also give thanks to the students and staff in the computing department and my friends who graciously consented to participate in the usability experiment of this study.

Last, but most importantly, my special thanks to my mother, who has always believed in me, without her love and support, I could not be here today.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

Nowadays, as users demand highly usable software [31], usability plays a significant role in the success of a software application. The objective of usability testing is to obtain an overall figure of merit that describes the usability level of the software [5]. The result of this test can help to improve the quality of products and increase the productivity of their users. Traditionally, user comments play a significant role in usability testing but they are too subjective and sometimes contradictive. If usability could be evaluated objectively with metrics, some resources and manpower can be reduced.

There are many different kinds of mobile devices in our daily life now, such as pagers, cellular phones and personal digital assistants (PDA). One of the bigger mobile phone manufacturers, Nokia, has announced that they have sold 50 million smartphones at the end of 2002 and predicted the number have been risen to 100 million in 2003 [12]. The new smartphones support Java 2 Platform, Micro Edition (J2ME) technology [29] which can support software applications executing in mobile device. It can be observed that users have been upgrading their cellular phones to smartphones which have stronger data processing capabilities. It can be foreseen that the demand of the software applications in smartphones will increase rapidly in the coming years because of the growing demand of users. Many software organizations are focusing on this new potential market because it will generate more business opportunities [27].

At present, despite the economic environment downturn of telecommunication industry, the third generation (3G) data service market will play a more significant role of generating new revenues for the industry [21] as it can foster many new business models

or add value for the existing business. Other than the investment of 3G network, the software application of these devices should be good in usability in order to increase the scope of users, who have different prior knowledge, and learning curve [15] for using mobile devices.

Some software applications not only operate on desktop computers or laptops, but also on some mobile computing machines, such as PDAs, two-way pagers and smartphones. The PDA has become more popular in last few years because of its small size and convenience for carrying. Many software development companies have already focused on this market because there is a high demand for different kinds of application which can take advantage of this high mobility environment. Some researches on the usability issue for PDA [3, 19, 20, 26, 33, 22] have already been reported.

The user requirement of the software application in smartphone is entirely different from that in desktop; and also the PDA users have different expectation of the software compared with those of smartphone users. However, there is a lack of study about smartphone, especially concerning about its software usability. At the same time, this platform has its own advantages and constraints, such as much smaller physical size for input and output. As a result, the software usability in this environment is usually not satisfactory. For example, users pay too much attention when using the application so that less focus on the situation when walking down the street. Thus, an investigation of the software usability problem in this platform is necessary.

The objective of this project is to develop a set of metrics that can be applied to evaluate the usability of the software applications in mobile phone. The existing software

usability metrics for desktop environment would be studied in order to obtain a comprehensive reference for further investigation in the mobile environment.

Law and Hvannberg [17] stated that the shared research context is not sufficient in the field of usability engineering. Researchers have proposed many usability metrics but these metrics use different measuring logic and output scales. As a result, they are confusing and difficult to use. This difficulty is compounded by the fact that some researchers provide only the name or concept of a metric without detailed descriptions or any computational method. Yet another problem is the lack of a comprehensive reference work identifying all the usability metrics. Consequently, it is not clear what usability metrics should be used in a given context and for a given objective. As a result, a standardized list of usability metrics will be of great value in usability evaluation.

However, before we constructed standardized usability metrics list, we had to collect most of the usability metrics and standardized them using a common scale. A consistent scale among the various metrics can enable developers to understand the usability of the application under evaluation. Thus, we constructed five formulae in order to convert different types of usability into the same scale.

Usability metrics may be applied under two types of evaluation: static and dynamic. Static evaluation means that the metric is computed without having execution of the application. Dynamic evaluation means that the metric requires the application to be executed. Usually, dynamic evaluation is not feasible for automation because its tasks require human interaction. Therefore, if the metric inputs involve users' feeling, checking or comment, they cannot be automated.

The usability evaluation is often done in the testing stage, a very late stage of development. If the developer finds a usability defect, the only solution may be to redesign the software [24], which could be very costly. Yet this need not be so, as different metrics can be applied at different development stages to measure different aspects of the usability of software. Thus, it is possible to detect some usability problems in the early stages of a product's development.

Other than evaluation type and phase of the development, the metrics were also classified based on their other characteristics, such as usability attributes and automation levels. The resulting usability metric list provides a useful reference for software developers who want to measure the usability of their products.

After the classified usability metric list had been constructed, the usability metrics which are appropriate for the software in mobile phone platform were selected according to the specific environment in this platform. In addition, some new usability metrics should be created because not all the usability requirement in this platform is covered by the metrics in the list. Our scope of the smartphone environment and the software type are described next.

There are many different types of smartphone in the market. Our research scope of smartphones is the most common type in the market at present which have reasonable prices. Also, their control and layout is very similar to previous GSM phones. This popular smartphone type let us know more about the common usage problems of the software application caused by less powerful processor, small memory size and limited input / output interface. Our scope will be described in detail in Chapter 4.

Apart from games, the e-commence application for retailing services is another common application in smartphones which can be used anywhere and anytime. Therefore, we have focused on this type of application in our research. According to Chan, Lee, Dillon and Chang's book [4], the e-commerce system for retailing should provide the following features:

1. Selection method for customers' purchase
2. Mechanism for order creation and submission
3. Appropriate mechanism for products / services distribution
4. Function for browsing the products / services by subject or category
5. Function for searching the products / services for particular good
6. Mechanism to provide customer service and obtain their feedback

Some of these features are not related to the software usability. For example, mechanism for goods distribution is related to the business model rather than the design of the application. Our usability evaluation for this kind of application has been conducted according to the above related features of 1, 2, 4 and 5.

My dissertation shows the research on the usability metrics consolidation and the usability issues of software application in smartphone. The structure of the dissertation is as follows. Firstly, some of the background information of this study and related research about the usability metrics and mobile devices are presented in Chapter 2. In Chapter 3, we describe the method of re-packaging the collected metrics, different groups of usability metrics and the dimensions of the classification. They are the criteria to construct the usability metric list. Next, the usability issues and problems which are caused by the limitations or constraints of smartphones are described in Chapter 4 and

then the hypotheses of this research and their adjustments after pilot test is discussed in Chapter 5. In Chapter 6, the experiment of our empirical study is described in detail, including the design of the experiment, its procedure and our hypotheses. Chapter 7 shows the findings of our analysis. This chapter suggests how to improve the usability of the software application in the smartphone platform and discuss the metrics which were applied to evaluate this software. Finally, our main conclusions are outlined along with some suggestion for future research in Chapter 8.

Chapter 2 Literature Review

In this chapter, firstly, the definitions of the usability, usability attributes and usability metrics are introduced. There are very important definitions for the whole research, because they are related to the goal setting and the experiment achievement for the project. Then some of the studies about mobile devices are presented in order to illustrate the limitations and usability issues in this environment. In fact, there are lack of usability research for software application of smartphone but we used the studies of other mobile devices, such as PDAs and WAP Phones, in order to obtain some comment usability issues with smartphone. Finally, some of the usability experiments in the mobile devices are referenced in order to design our usability evaluation of the software application in smartphone.

2.1 Usability Attribute and Metrics

Usability is "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" [24]. It can be observed that usability consists of three attributes, which are effectiveness, efficiency and satisfaction [10].

- Effectiveness is defined as "the accuracy and completeness with which users achieve certain goals".
- Efficiency is defined as "the relation between the accuracy and completeness with which users achieve certain goals and the resources expended in achieving them".
- Satisfaction is defined as "the users' comfort with and positive attitudes towards the use of the system".

According to Frøkjær, Hertzum and Hornbæk [10], the quality of solution is an indicator of effectiveness; the task completion time and learning time are the indicators of efficiency; and user preferences are indicators of satisfaction. In Chapter Three, the relationship between metrics and usability attributes will be determined based on the above criteria.

These three usability attributes usually apply to the WIMP (windows, icons, pointer and mouse) software [13] or web application in previous research. Our study was also based on these attributes to evaluate the usability of software application in smartphone. Fig. 2.1 illustrates the relationship between the overall usability level and the usability attributes.

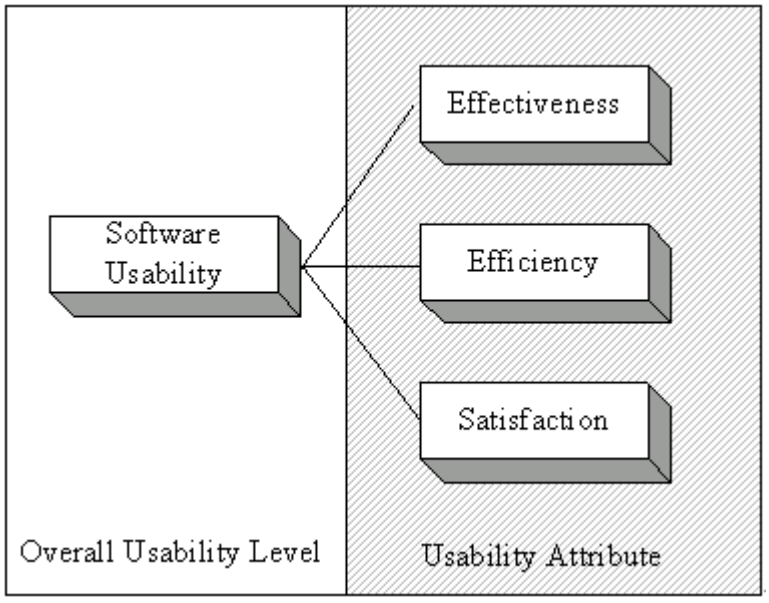


Fig. 2.1 The Relationship between Overall Usability Level and Usability Attributes

Constantine and Lockwood [7] defined usability metric as "quantitative measures of the quality of a system in terms of its actual or estimated usability". In usability evaluation,

many raw data can serve as inputs of the metrics, such as the number of key input or response time. The value of the metric is derived from a numerical computation, which is an objective method to measure and compare the usability.

The above research gave us not only the definition of software usability, but also that of effectiveness, efficiency and satisfaction. All these definitions are the core rules in this research, not only in the experiment design but also in the hypothesis development. The following paragraphs show some research about usability metrics.

There are some existing studies of usability metrics which target on the software application on desktop computer. Ivory, Sinha and Hearst [14] proposed a set of metrics which could be used to check the usability of web pages. They, however, provided only the metric names along with a very brief description. Babiker, Fujihara and Boyle [1] also presented a set of metrics for web pages, accompanied by detailed explanations and formulae. Noble and Constantine [22] proposed three metrics for user interface design: Layout Uniformity, Task Concordance, and Visual Coherence. ISO 9126 [9] had a list of usability metrics which can evaluate the WIMP environment. Dix's book provided a name list of usability metrics and some idea on how to build standardized usability metrics [8].

Although the metrics in above studies are focused on the desktop computer, some of those which are appropriate for mobile phone environment will be re-packaged for applications in smartphone. In addition, there are no standards for usability metrics, and different researches are using different methods to calculate for usability metrics results. At the same time, there is no common way for the definition of those metrics results. As

a result, there is a gap for this research to define a common scale for the metric result value in order to let the end-user know what constitutes a good result.

2.2 Mobile Devices Environment

White's paper [29] mentioned some background information and the structure of J2ME technology. Helal [11] identified some useful information related to the software usability for the device which supported J2ME, such as the memory size, the computational power of the processor, and the connectivity of the network. This information showed some of the usability issues of software application in the mobile phone environment. Our research focused on this kind of mobile phones, which make use of the J2ME technology, with the software developed in Java language.

Weiss's book [28] classified handheld devices into 3 different groups: PDA, mobile phone and two-way pager. If a device had the usability characteristics of both the mobile phone and PDA, it should be classified as communicator. This book clearly illustrated not only the common problems of these mobile devices, but also their different usages, different constraints for input / output and interface limitation for our analysis.

Among the different groups of mobile device, there are many well studied research about PDA. Kristoffersen and Ljungberg [16] investigated the relation between user interface design and ease of use in handheld Computer Supported Cooperative Work (CSCW). Although their research focused on PDA, some of their findings can be applied to other types of mobile devices. Applications running on mobile devices had the following key properties:

1. The external tasks which have to be performed alongside mobile devices are more important compared with the tasks operated in desktop computer in the office.

2. In the mobile environment, users' hands are usually occupied. For example, they need to manipulate physical object during their work. However, the office staff can place both of their hands on a keyboard.

3. Users on the move need to play a high level of visual attention to inspect the environment around them but the office staff can pay a large degree of visual attention towards a computer.

4. Users need to complete this kind of task in high mobility, as opposed to the office staff who can stay in their desk to finish their tasks

The above key properties gave us a good reference when we were defining the usability metrics, research hypotheses and designing the usability experiment because the above usability issues are valid in the smartphone environment.

Another study investigated how to build multi-data service in GPRS infrastructure [21]. In this research, the researchers used smartphone to present three demonstrations about data representation which bring out some usability issues. They used the number of key pressings in a task to compare the usability level of each of the system designs. Voice entry was suggested for character string input in the software application because there are many usability problems using keypad on smartphone. They hoped this new input system, voice entry, could improve the usability of the application.

The 3D audio pie menu [2] was introduced based on gesture input and non-speech audio output for wearable computing. The head gestures together with egocentric sounds were a successful interaction technique which provided a significant improvement in

usability. Because of the ‘eye-free’ control platform, this technique had less impact on walking speed. Thus, the sound effect from smartphone should play an important role for the usability of the application in mobile environment.

Masoodian and Lane [18] suggested that small physical size enhances the portability of mobile phones in order to increase their popularity. However, this also decreased the capabilities of input and output of the phones as their interfaces (screen and keypad) were limited by their small sizes. At present, mobile phones usually have many menu options because of their multitude of functions which reduce their usability. According to Masoodian and Lane’s experiment, textual style interface was more effective than graphical style, as WAP-based mobile phones had some limitations for accessing graphical information.

According to the above studies, also, our own observation of the software application in the smartphone platform, the limitations and the constraints of smartphone are systemically presented in Chapter 4. Also, some of the usability metrics, such as “Key Input Number”, “Single / Two Hand Control” and “Sound Effect”, are proposed in this chapter according to the reference of the above studies. In Chapter 5, some hypotheses, which are related to the software application in the mobile environment, are proposed by the information to the previous studies.

2.3 Usability Evaluation of Mobile Device

This section introduces some previous experiments which were used to evaluate the usability of mobile devices. They are the good references in our research because we

should design a usability test in order to qualify our proposed usability metrics and hypotheses for the software application in smartphone platform.

Kiili carried out a case study of the usability in WAP phone for mLearning [15]. He identified some obstacles of the user interface in some mobile phone models. Although there are some differences of the interface between WAP phone and smartphone, they have some common problems, such as limited input and output interfaces. This paper also gave a method of usability evaluation of WAP phone and provides a good reference for doing a usability evaluation of smartphone.

Pirhonen, Brewster and Holguin [23] tried to use gesture and non-speech audio towards a mobile software application, a music player, in PDA. This could improve the performance when user was interacting with the device because he could control the device without looking at it. User only needed to use his / her finger to sweep across the PDA screen for input and non-speech sound was then output to user. There were two parts in their experiment: formal usability experiment and video analysis. Their procedures, such as phone usage briefing session, using mobile device during walking and video capturing, were referenced for the design of our experiment.

Wigdor and Balakrishnan [30] proposed a new input strategy in which users tilt the mobile devices. For the traditional Triple-tap test input in mobile phone, users should repeatedly press the same key to enter one character. By tilting the mobile phone with appropriate direction, only one key press was needed for a character input. They used t-test for the data analysis and concluded that their new input strategy had significant improvement compared with traditional strategy.

Some techniques in the above studies, such as “Thinking Aloud”, video analysis and the run-down of the whole experiment were referred to during the design of our own usability experiment. The detail of this experiment is shown in Chapter 6.

Chapter 3 Constructing a Usability Metric List

This section presents the four key steps for consolidating and constructing a usability metric list: collection, consolidation, normalization, and conversion (Fig 3.1). In this metric list, all the usability metric result values will have the same scale of output, 0 to 100, and the usability level is directly proportional to this value.

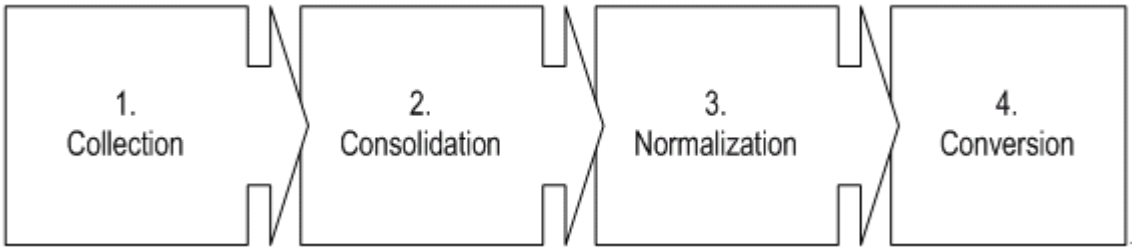


Fig. 3.1 Four Steps to Construct a Usability Metric List

3.1 Collection of Metrics

Metrics from different publications are often inconsistent. Some metrics have very detailed descriptions while others are simply named. Thus, the first step to construct a usability metric list is to collect all metric names from different sources. Metrics are then grouped according to whether or not they are similar in input, output or name.

3.2 Consolidation of Metrics

After grouping the metrics based on several dimensions, such as evaluation type and usability attribute, redundant metrics with overlapping properties are eliminated. Some metrics have different names but are essentially the same; for example: “Number of available commands not evoked” and “Evident Function” measure similar aspects. Some metrics have similar inputs but different calculation methods, outputs and scales.

The metrics, such as “Frequency of help and documentation” and “Effectiveness of help system”, may be used to measure the same concept. Some metrics *include* others; for example, “Number of times the interface misleads the user” can be represented by “Function Understandability”. Some usability metrics can therefore be combined to eliminate overlapping metrics.

3.3 Normalization

The next step is the normalization of metrics. As our study mainly focuses on standardizing usability metrics, it is very important to choose a suitable scale. We have chosen the 100-based scale. The reason for selecting this scale is that many people are familiar with it and it is similar to the percentage system, except that our scale will not be greater than hundred or smaller than zero. According to this 100-base scale, when the metric result is close to 100, the usability level is “Good”. At the same time, when this value is nearly 0, the usability level is “Bad”. This scale is understandable even for users who are not experts in usability.

3.4 Metric Conversion

The last step is to convert the metrics to a 100-based scale. After conversion, the names of some metrics may no longer be suitable; for example, the name “Text Positioning Count” should be changed to “Text Positioning”.

Some metrics such as “Layout Uniformity”, “Task Visibility”, and “Visual Coherence” are already in the 100-based scale and do not need any conversion.

For the purpose of our study, five types of metric have been identified. They are shown in Table 3.1. Column 1 lists the names of metric type. Column 2 provides a brief description of the metric type. Column 3 lists the required conversion.

Metric Type	Characteristic	Conversion
Bad Count	Usability metrics are based on the number counts of some bad situations.	Deduce Base
Portion	Usability metrics have two inputs and the first input is a portion of the second input.	Portion
Uniform Output	Usability metrics have uniform scaling for their outputs and their output features are very similar to the 100-base scaling.	Scaling
Reference Value	Usability metrics are based on a comparison with a Reference Value. The closer the measured data to the Reference Value, the higher is the usability.	Reference Dependent
Standard Value	Usability metrics are based on a comparison with a Standard Value, which will be the usability will be considered the highest.	Standard Dependent

Table 3.1 Metrics Type List

Metrics belonging to the type “Bad Count” are based on the count of some bad situations such as “Number of errors” and “Number of times user loses control of the system”. If the count is large, the usability level of the application is low.

Metrics of the type “Portion” requires two inputs and the first input is a portion of the second input. The usability level of the application is high if the values of the first input and second input are close to each other.

Metrics of the type “Uniform Output” have uniform scaling of their outputs that are not 100-based.

Metrics of the type “Reference Value” are used mainly for evaluating web pages, such as “Word Count”, “Link Count” and “Page Size”. They use a “Reference Value” for comparison. The Reference Value acts as the ideal value for the metric. The usability levels of the application are high when the measured data is close to the Reference Value. The reference value can be obtained from statistics or surveys [14]. As an

example, a web site may have different targeted user groups, such as children, students and business administrators. Different users will have different preferences for web pages; for example, children prefer colourful web pages with many pictures while business administrators prefer formal and informative web pages. Using suitable reference values, the usability level among different target groups can be determined accurately.

Metrics of the type "Standard Value" are used mainly for evaluating windows applications which involve WIMP [13]. They use a Standard Value for comparison. Standard Values can be obtained from previous testing results or software engineering standards. There are two types of Standard Value: Standard Maximum (SMAX) and Standard Minimum (SMIN). For some cases, SMAX represents the perfect value and the ideal case is when the measured data is below SMAX; for other cases, SMIN represents the perfect value and the ideal case is when the measured data is above SMIN.

Five ways of re-computing the metrics have been identified. Let RV be the Reference Value, SV be the Standard Value, IV be the Input Value which represents the value of the metric before conversion, UB be the Upper Bound, and LB be the Lower Bound.

3.4.1 Deduce Base Conversion (DBC)

The formula below shows the calculation of Deduce Base Conversion.

$$\mathrm{DBC}_{\mathrm{Output}} = \begin{cases} 100 - (\alpha \times IV) & , \text{if } \alpha \times IV < 100 \\ 0 & , \text{if } \alpha \times IV \geq 100 \end{cases}$$

where α is a deduction parameter. α serves as a weighting factor for each "bad" event. For example, for "Total Error", α may be set at 5, meaning that the output will be reduced to 0 if 20 errors have been detected.

The second part of the formula is needed to ensure that no negative output values are produced. If the IV is small, the output value will tend to one hundred, meaning that the application does not have many bad situations such as errors. However, if the IV is very large, the output will become zero, meaning that the usability is not good.

3.4.2 Portion Conversion (PC)

The formula below shows the calculation of the Portion Conversion.

$$\mathrm{PC}_{\mathrm{Output}} = 100 \times \frac{IV}{TIV}$$

where TIV is the Total Input Value.

As IV is a portion of the TIV, the usability level of the application is high if their values are close to each other. For example, the metric "Evident Functions" is equal to "Number of functions identified by the user" over "Total number of actual functions".

3.4.3 Scaling Conversion (SC)

Scaling Conversion is based on a simple mapping method that maps the original scale to the 100-based scale. The conversion formula is shown below:

$$\mathrm{SC}_{\mathrm{Output}} = 100 \times \frac{IV - LB}{UB - LB}$$

For example, the output range of Task Concordance [22] is from –100 (LB) to 100 (UB). If the value of Task Concordance before conversion is –10, the new metric value after the conversion is:

$$SC_{Output} = 100 \times \frac{-10-(-100)}{100-(-100)} = 45$$

3.4.4 Reference Dependent Conversion (RDC)

In Reference Dependent Conversion, the original metric is converted into a value relative to the RV. There are two adjustment parameters for this conversion, the UB and the LB. They are the boundaries of the acceptable range of IV. The output of RDC is based on comparing the ratio between the difference of RV and IV and the difference of boundary (UB or LB) and IV. If IV (i.e. word count) is close to the RV (i.e. reference word count of a target group), the output of the metric will tend to a hundred.

In Fig. 3.2, if IV reaches or becomes smaller than LB, or reaches or becomes greater than UB, the output will return 0. The output will be larger than zero if the IV is between LB and UB. If the IV is close to RV, the output will tend to one hundred.

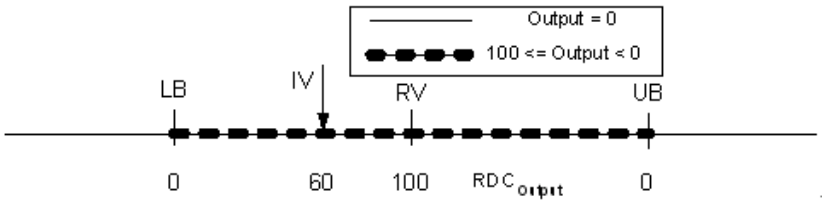


Fig. 3.2 RDC

For example, consider the metric "Word Count" and assume its RV is 5000 words. Let LB be 4000 words and UB be 7000 words. The acceptance range of the IV is therefore between 4000 and 7000 words and the output is 100 if IV is equal to RV. The metric

value will be 50 if IV is 6000 or 4500, which is the mid value of the lower acceptance interval (from LB to RV) or upper acceptance interval (from RV to UB) respectively. The conversion formula is shown below:

$$\mathrm{RDC}_{\mathrm{Output}} = \begin{cases} 0 & , \text{if } IV < LB \text{ OR IV} > UB \\ 100 \times \left(1 - \dfrac{RV - IV}{RV - LB}\right) & , \text{if L}B \ \leq IV < RV \\ 100 \times \left(1 - \dfrac{IV - RV}{UB - RV}\right) & , \text{if } RV < IV \leq UB \end{cases}$$

3.4.5 Standard Dependent Conversion (SDC)

The Standard Dependent Conversion requires Standard Values (SV) for calculation. There are two cases: SMAX and SMIN. Fig. 3.3 illustrates the SMAX case. The acceptable range of this conversion is smaller than UB. Therefore, the output will reach zero if IV ≥ UB. If IV < SMAX, the output will become a hundred. The larger the IV when compared to SMAX, the smaller the output.

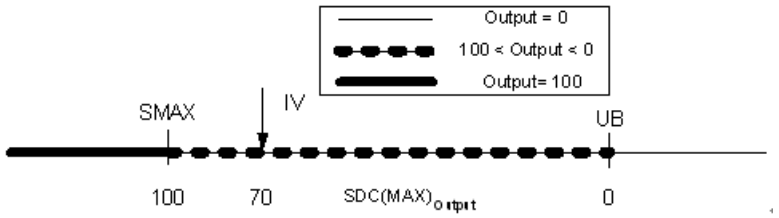


Fig. 3.3 SDC with SMAX

Similarly, the opposite case for SMIN is shown in Fig. 3.4.

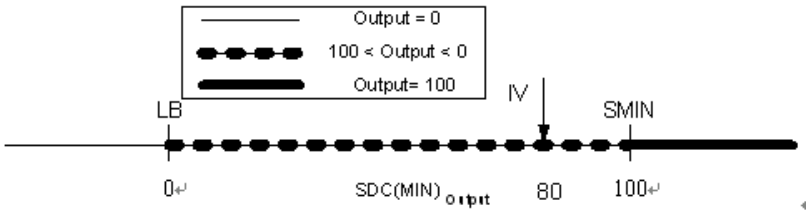


Fig. 3.4 SDC with SMIN

The metric "User Expression Ratio Level" is used as an example. It should use SDC(MIN) because its IV is the user expression ratio and the larger this ratio, the better the result. Let SMIN be 5, and LB be 4. The acceptance range of IV is therefore above 4 and the output is 100 if IV ≥ SMIN. The output of the metric is 50 if IV is 4.5 which is the mid-value of the interval between LB and SMIN. The standard dependent conversion formulae are shown below:

$$\mathrm{SDC(MAX)}_{\mathrm{Output}} = \begin{cases} 100 & , \text{if } IV \le SV \\ 100 \times \left(1 - \dfrac{IV - SV}{UB - SV}\right) & , \text{if } SV < IV \le UB \\ 0 & , \text{if } IV > UB \end{cases}$$

$$\mathrm{SDC}(MIN)_{\mathrm{Output}} = \begin{cases} 100 & , \text{if } IV \ge SV \\ 100 \times \left(1 - \dfrac{SV - IV}{SV - LB}\right) & , \text{if } LB < IV \le SV \\ 0 & , \text{if } IV < LB \end{cases}$$

3.5 Metrics Classification

Metrics can be classified along four key dimensions: evaluation type, applicable development stage, measured usability attribute, and level of automation.

3.5.1 Evaluation Type

There are two evaluation methods, static and dynamic, for usability metrics. The static metrics can be computed without any execution of the application. The dynamic metrics should be calculated during the execution of the application, either by the users or automatic tools.

3.5.2 Development Stages

Metrics can be applied in three applicable stages, namely, Prototype Stage, Coding Stage and Execution Stage. Some metrics are applicable to more than one stage.

3.5.2.1 Prototype Stage

The prototype stage occurs after requirement gathering and system design but before implementation. Developers often implement interface and function prototype according to the analysis from the system design stage. Metrics that can be applied at this early stage will act as guidance for developers to understand the usability level of their user interface and functional design. Developers can then modify the software design to enhance its usability as early as possible, so as to avoid a high rework cost.

3.5.2.2 Coding Stage

The coding stage occurs after the prototype stage and before all functions of the application are completed. At this stage, the developers can understand some aspects of usability by studying the source code. Metrics available at this stage are mainly for checking whether or not the code contains any careless mistakes. Mistakes appearing in the code will affect the usability level of the software, such as “Consistency of Function Wording”. Some metrics that relate to the layout representation can also be determined at this stage.

3.5.2.3 Execution Stage

The execution stage involves System Test and User Acceptance Test. Since most of the functions have been completed at this stage, the usability of the completed application can be evaluated dynamically. Metrics used in this stage will include feelings of and comments from users. An example metric is "Function Understandability". This kind of information cannot be obtained at previous stages.

3.5.3 Usability Attributes

As discussed earlier, there are three usability attributes: effectiveness, efficiency and satisfaction. Metrics may be classified based on the particular attribute that they measure.

If a metric measures the completeness of the task, it will be mapped to the effectiveness attribute. This type of metric focuses on the accuracy of the task, meaning whether or not the task can be completed and whether or not the outcome is consistent with users' expectations. "Total Error" and "Success Failure Ratio" are metrics under this attribute as they measure the accuracy of the task.

If a metric measures some time-related elements, such as completion time or learning time of each task, it will be mapped to the efficiency attribute. "Number of steps", "Number of command uses", "User operational frequency" and "Efficiency of task completion" are metrics under the efficiency attribute because the inputs of these metrics contain time elements. "Tab Sequence in One Page" also belongs to this attribute as it shows whether users can perform the task smoothly, which affects the task completion time.

If a metric measures the feelings of a user, such as user preference or user understanding, it will be mapped to the satisfaction attribute. This type of metric is based on the subjective opinion of users. "Function Understandability" and "Function Favorable" are the metrics under this attribute.

3.5.4 Automation

Usability metrics may be automatically computed at four levels.

• **Interface Review Automation:** The layout of the software application should be designed in the early part of development. If the layout is specified in the document, the usability metrics cannot be calculated automatically. However, when the templates of the interface have already been developed, automatic tools can capture the raw data from the source code of the interface template and this data can become the input of some metrics.

• **Source Code Automation:** Some metrics, which do not require any human decision, ability, or feelings, can be computed automatically based on the source code of the software application. For example, some usability defects can be discovered by examining the source code. One of the advantages of source code automation is that these metrics be automated easily as the source code can be directly evaluated by testing tools.

• **Test Script Automation:** Usability testing is always based on test scripts. During preparation for usability testing, usability experts prepared some test scripts based on the usage scenario of the software application. Some simulation tools can execute the application automatically according to the test script and then the usability metric can be

captured by those tools. This type of automatic usability test tools is similar to the automatic regression test tools.

• **Product Automation:** Product Automation is quite similar with Test Script Automation. It requires not only a completed product, but also users' execution. An automatic tool can capture some raw data, such as execution time, during software testing by users.

Table 3.2 and 3.3 show the metrics belonging to "Static Analysis" and "Dynamic Analysis" respectively. They also list those metrics with other different classifications. Column 1 provides the abbreviations of the metrics, and the symbol below the abbreviations shows the type of conversion applied. Column 2 lists the metric names. Column 3 lists the raw data, which is essentially the conversion input. Column 4 provides extra information about the metrics. Columns 5 to 7 show the list for the usability attributes classification ("E" for "Effectiveness", "F" for "Efficiency", and "S" for "Satisfaction"). In these three columns, "X" indicates that the metric belongs to the attribute. Columns 8 to 10 show the list for the Development Stages classification ("PS" for "Prototype Stage", "CS" for "Coding Stage" and "ES" for "Execution Stage"). In these three columns, if any of these columns is marked by a character ("I" for "Interview Review Automation", "D" for "Source Code Automation", "T" for "Test Script Automation", "P" for "Product Automation" and "N" for "No automation"), that metric can be applied in that development stage.

M. Abb.	Metric Name	Formula Input	Remarks	U. Attributes			D. Stage		
				E	F	S	PS	CS	ES
	Static Analysis								
AL ▲	Access Layer	IV: Number of layers to access the function	1. Total steps needed to perform before actually accessing the function 2. Fewer steps are preferred		X		N	I	I
CFW ▲	Consistency of Function Wording	IV: Number of inconsistent function wordings	The wording of the same function displayed in different places should be consistent, otherwise, user can easily be confused	X				D	D
TS ▲	Tab Sequence	IV: Number of nonsequential tab orders in a page	Tab sequence should follow the order of using objects in interface. User will find nonsequential tab inconvenient		X		N	I	I
BTPI ●	Body Text Proportion Index	IV: Number of words that are body text / Number of words that are display text RV: Reference body text proportion	1. This determines the proportion between the number of words that are body text and the number of words that are display text (i.e. headers) 2. Users respond differently to page with different body text proportions			X		D	D
EBTPI ●	Emphasized Body Text Proportion Index	IV: Body text that is emphasized / Number of words that are body text RV: Reference emphasized body text proportion	1. This proportion deals with the amount of body text that is emphasized(e.g., bold, capitalized or words that are close to exclamation marks) 2. Users respond differently to pages with different emphasized body text proportions			X		D	D
GSP ●	Graphic Size Proportion	IV: Bytes of graphic in a page / Total bytes in the page RV: Reference proportion of bytes of graphics in a page	1. Portion of bytes in a page that are used for graphics 2. More graphics will please users from some target groups but this will affect the response time		X			D	D
PS ●	Page Size	IV: Total bytes for the page RV: Reference bytes size in a page	1. Determines the total bytes in a page as well as elements graphic and stylesheets 2. Page size affects the response time which consequently affects the emotions of users		X			D	D
TCL ●	Total Color	IV: Total number of colours employed RV: Reference number of colours employed	Determines the total number of colors in a page and whether users are comfortable with that			X		D	D
TCT ●	Text Cluster	IV: Number of text areas RV: Reference number of text areas	1. Text areas highlighted with color, bordered regions, rules or lists 2. Determines whether users are comfortable with the number of colored text areas			X		D	D
TF ●	Total Font	IV: Total number of fonts employed RV: Reference number of fonts employed	1. Total number of fonts employed (i.e, face + size + bold + italic) 2. Determines the total number of fonts in a page and whether users are comfortable with that			X		D	D
TG ●	Total Graphic	IV: Number of graphics on a page RV: Reference number of graphics	1. Total number of graphics, not including graphics specified in scripts, applets and objects 2. Determines the total number of graphic in a page and whether users are comfortable with that			X		D	D
\TL ●	Total Link	IV: Total links on a page RV: Reference number of links	Determines whether or not the number of hyper-links in a page is appropriate	X				D	D
TP ●	Text Positioning	IV: Number of changes in text position RV: Reference number of changes in text position	1. Changes in text position from flush left 2. Determines whether users are comfortable with the number of changes in text			X		D	D
TW ●	Total Word	IV: Total number of words contained in a page RV: Reference number of words	Determines the total number of words contained in a page and whether or not users find that comfortable			X		D	D
LU	Layout Uniformity	N.A.	Examines the design of the interface and determines whether or not it is good according to the designer's viewpoint [7]			X	N	I	I
TV	Task Visibility	N.A.	Discovers the relationship between visibility of features and necessary capabilities to finish tasks [7]	X			N	I	I
VC	Visual Coherence	N.A.	Examines the interface layout and determines the method to grouping related objects together and separating non-related objects [6]			X	N	I	I

Table 3.2 Usability Metrics List under Static Analysis

M. Abb.	Metric Name	Formula Input	Remarks	U. Attributes			D. Stage		
				E	F	S	PS	CS	ES
	Dynamic Analysis								
CFU ♦	Completeness of Function Understood	RI: Number of functions understood TI: Total number of functions	Determines the amount of functions users feel they understand in comparison with the actual number of functions in the program			X			N
EE ♦	Efficiency on Error	RI: Time to complete a task - Time spent in error TI: Time to complete a task	Determines the amount of time not spent on error in comparison with the total amount of time to complete the task		X				P
EF ♦	Evident Functions	RI: Number of functions identified by the user TI: Total number of actual functions	Determines the amount of functions users can identify in comparison with the total number of functions in the program	X					N
EFHS ♦	Effectiveness of Help System	RI: Number of tasks successfully completed after accessing online help TI: Number of tasks tested accessing online help	Determines the amount of tasks completed successfully after using online help in comparison with the amount of tasks tested	X					P
EFTC ♦	Effectiveness of Task Completion	RI: Number of completed tasks TI: Total number of tasks	Determines the number of tasks which can be completed in comparison with the total number of tasks	X					T
EUA ♦	Error Undo-ability	RI: Number of functions successful customized TI: Number of error conditions tested	Determine how many functions which are customized successfully because of the found errors in comparison with total error conditions	X					N
EUH ♦	Easy of Use Help System	RI: Number of tasks having correct online help TI: Number of tasks tested for which have online help	1. Determines the amount of correct online help 2. User will find it is easy to use the application if the online help is correct	X					N
FF ♦	Function Favorable	RI: Number of functions having favorable comments from users TI: Number of functions available	Determines the amount of functions user are favorable of in comparison with the total number of functions available			X			N
FU ♦	Function Understandability	RI: Number of interface functions, the purpose of which is correctly described by the user TI: Number of functions available from the interface	Determines the amount of interface functions purpose users can describe correctly in comparison with the total number of interface functions in the program			X			N
IAC ♦	Interface Appearance Customizability	RI: Number of turns which user successes to select input or output expression TI: Number of turns which user tried to select input/output expression	Determines the portion of users successfully select input or output expression	X					P
IUA ♦	Input Undo-ability	RI: Number of input error which the user successfully corrects TI: Number of attempts to correct input errors	Determines the number of input errors which users successfully corrects in comparison with the total number of input error which users attempted to correct	X					P
SEM ♦	Self Explanatory Error Message	RI: Number of erroneous conditions which users can propose the correct recovery action TI: Number of error conditions tested	Determines the amount of errors which users can propose the correct recovery action once the system provide the instruction	X					N
DT ▲	Disrupt from Task	IV: Number of times the user is disrupted from a work task	The higher the number of times users being disrupted, the more marks deducted.	X					N
LC ▲	Loss Control	IV: Number of times the user loses control of the system	The higher the number of times users loose control, the more marks deducted.	X					N
MU ▲	Misleading Users	IV: Number of times the interface misleads the user	The higher the number of times the interface misleads the user, the more mark deducted.	X					N
RF ▲	Repetition of Failure	IV: Number of repetitions of failed commands	The higher the number of repetitions of failed commands, the more marks deducted.	X					P
TE ▲	Total Error	IV: Number of errors	The more errors, the more marks deducted.	X					T
WAP ▲	Work Around Problem	IV: Number of times the users need to work around a problem	The higher the number of times users have to work around a problem, the more marks deducted.	X					N
EL ■	Easy of Learning	IV: Mean time take to learn to use a function SV: Standard time to learn to use the function	1. Compares the learning time of a function with the standard learning time 2. The shorter the learning time, the easer the function can be learnt			X			N
ECHS ■	Efficiency of Help System	IV: Time spent using help SV: Standard time spent using help	The help system is not efficient enough if users spend a long time on it		X				P
ECI ■	Error Correction Index	IV: The end time when specified type errors are completely corrected – The start time of correcting specified type errors SV: Standard time for error correction	Obtains the amount of time needed to correct specified type of errors in comparison with standard error correction time		X				P

ECTC ■	Efficiency of Task Completion	IV: Time to complete a task SV: Standard time to complete a task	The efficiency of a task is low if the task completion time is much greater than the standard time		X				T
UC ■	Used Commands	IV: Number of commands used SV: Standard number of commands used	It is not good enough if too many commands are required to finish a task in comparison with the standard number		X				T
CL □	Customizability Level	IV: Number of attempts to customize SV: Standard number of attempts to customize	If the number of attempts to customize is greater than standard number of attempts to customize, full marks will be obtained			X			N
EFTCU □	Effectiveness of Task Completion in Unit Time	IV: Number of completed tasks in unit time SV: Standard number of completed task in unit time	1. Determines the number of completed tasks in unit time in comparison with the standard number 2. If the number of completed tasks in unit time is greater than standard number, full marks will be obtained		X				T
RB □	Regressive Behaviours	IV: Number of regressive behaviors SV: Standard number of regressive behaviors	1. Determines the number of regressive behaviors in comparison with the standard number 2. More regressive behaviors represents easier learning of the software	X					P
RFRI □	Recalled Features Ratio Index	IV: Number of good features recalled by users / Number of bad feature recalled by users SV: Standard recalled feature ratio	The recalled features ratio will be found and compared with standard ratio in order to obtain the index			X			N
SFRI □	Success Failure Ratio Index	IV: Number of successful tasks / Number of fail tasks SV: Standard success failure ratio	The success failure ratio will be found and compared with the standard ratio in order to obtain the index	X					T
UERI □	User Expression Ratio Index	IV: Number of the user expresses satisfaction / Number of users expresses frustration SV: Standard user expression ratio	The user expression ratio will be found and compared with standard ratio in order to obtain the index			X			N
UOFI □	User Operational Frequency Index	IV: Number of turns which users use the specific software functions / Operation time SV: Standard user operational frequency	1. The user operational frequency will be found and compared with the standard frequency in order to obtain the index 2. If the user operational frequency is greater than standard frequency, full marks will be obtained			X			T
TCC ○	Task Concordance	IV: Output value of the Task Concordance metric before conversion	Examines the relationship between the frequency and the difficulties of a given task [22].		X				T

●: Reference Dependent Conversion ◆: Portion Conversion ▲: Deduce Base Conversion
■: Standard Dependent Conversion (Max type) □: Standard Dependent Conversion (Min type) ○: Scaling Conversion

Key: E: Effectiveness F: Efficient S: Satisfaction
PS: Prototype Stage CS: Coding Stage ES: Execution Stage
I: Interface Review D: Source Code Automation T: Test Script Automation
P: Product Automation N: This metric can apply in this stage but no automation

Table 3.3 Usability Metrics List under Dynamic Analysis

3.6 Summary of Classifications

Table 3.4 summarizes the metric distribution. Column 1 lists the name of the development stage. Column 2 provides the abbreviations of the usability attributes ("E" for "Effectiveness", "F" for "Efficiency" and "S" for "Satisfaction"). Column 3 lists the number of metrics which can apply Interface Review Automation (I). Column 4 lists the number of metrics which can apply Source Code Automation (D). These two automation types belong to Static Evaluation. Column 5 lists the number of metrics which can apply Test Script Automation (T). Column 6 lists the number of metrics

which can apply Product Automation (P). These two automation types belong to Dynamic Evaluation. Column 7 lists the number of metrics which cannot be automated. In each row, there are three sub-rows and they represent the number of metrics belonging to a particular usability attribute. Column 8 lists the sub total of metrics which belong to an attribute and can be applied in a development stage. Column 9 lists the total number of metrics which can be applied in each development stage.

Dev. Stage	***Usability Attribute***	***Static Evaluation***		***Dynamic Evaluation***		***No Auto.***	***Sub. Total***	***Total***
		I	***D***	***T***	***P***			
Prototype Stage	E	0				1	1	5
	F	0				2	2	
	S	0				2	2	
Coding Stage	E	1	2			0	3	17
	F	2	2			0	4	
	S	2	8			0	10	
Execution Stage	E	1	2	3	5	8	19	48
	F	2	2	4	3	0	11	
	S	2	8	1	0	7	18	

Table 3.4 Distribution of the Metrics among Different Dimensions

3.6.1 Evaluation Type VS Development Stage

In the Prototype Stage and Coding Stage, the product is not completed and is not fully executable. Metrics based on Static Evaluation can be applied only in these two stages.

In the Execution Stage, the development of the product is complete and some metric data can be obtained from execution. The metrics which are based on Dynamic Evaluation can be applied at this stage. In fact, the metrics which at previous stages can be applied using static evaluation can also be applied at this stage.

3.6.2 Development Stage VS Automation Level

All metrics in the Coding Stage can be automated, but no metrics can be in the Prototype Stage. This is because at the prototype stage, the design of the product has only been completed in paper form and in this situation no automation can be applied. The metrics in the Coding Stage, however, can be derived from the source code. These metrics can be computed automatically by tools which can scan the code and collect the raw data for the metrics. Some metrics in the Execution Stage cannot be automated as they require users' comment and feeling.

At the Prototype Stage, we have identified five metrics, none of which can be automated.

Automation can be applied at the Coding stage with both Interface Review Automation and Source Code Automation. Although a completed product is not available, the source code should be available and can be evaluated with automatic tools to obtain the metric data.

The metrics for the Execution Stage may be applied by Test Script Automation and Product Automation because the product is completed and executable. Those metrics which do not require any user-feedback can be automated using the test scripts to simulate the usage scenarios. Some metrics which measure the ability of the user to control and operate software can also be automated by tracking the user interaction.

3.6.3 Usability Attribute VS Automation Level

In Table 3.4 we can see that only slightly more than half of the metrics (eleven out of nineteen) for the effectiveness attribute can be automated. This is because some of the metrics for the effectiveness attribute examine the completeness of tasks, which rely partly on the ability of the users, which is not easy to measure.

All the metrics for the efficiency attribute can be automated. This is because the metrics in this group contain a time element or are related to the finish time of tasks and can be captured by tools.

More than half of metrics (eleven out of eighteen) for the satisfaction attribute can be automated. Most of the automated metrics of this attribute are obtained by Source Code Automation. Some of the metrics for the satisfaction attribute require the information about a user's feelings, along with comments, and that kind of metric cannot be automated.

Chapter 4 The Criteria of Software Usability in Smartphone

In this chapter, first of all, the software usability issue in smartphone environment will be discussed. The usability issues in this environment are quite new and lack of research until now. Then our proposed usability metrics which are suitable for this environment is presented and a summary is concluded at the end of this chapter.

4.1 Common Usability Issues for Applications on Smartphone

First of all, the detail of the smartphone environment will be described. Then the scope of this research is defined according to the functionality of the hardware. Hardware comparison among desktop computing, PDA and smartphone will then be done. Finally, some usability issues of the software within the scope will be raised and discussed.

4.1.1 The Environment of Smartphone

Smartphones are similar to the traditional GSM smartphones which contain a keypad for the input system. Apart from the keypad, smartphone has two soft keys which are located under the phone screen. The functions of these two keys will change dynamically depending on the software design. Their functions will be labeled in the bottom of the phone screen. Between the two soft keys, there are direction keys which are used to control the cursor in the software application. Some smartphones also have a Select Button which has the same function as the Enter Key in a computer. Fig. 4.1 shows the common layout of the smartphone.

Fig. 4.1 The Layout of a Typical Smartphone

The computational power of the processor, 16- or 32-bit processor, is very slow [11] compared with the desktop computer or even PDA. The size of the memory in the smartphone is about 160 kilobytes to 512 kilobytes [11] and its capacity is quite limited. The speed of the network is not fast and it is a low-bandwidth wireless network [11] and is usually unstable.

Most of the smartphones have colour screens but they are quite small and the resolution is not high compared with the desktop monitors. Because of their high mobility, these

phones should be designed as small as possible. Smartphones can generate some sound effects which enhance the interaction with users.

4.1.2 Scope of Research and the Comparison among Different Environment

This study will focus on the usability problem of software in smartphone environment. In fact, there are three types of handheld devices which are smartphone, personal digital assistant (PDA) and two-way pager [28]. For the mobile machines which have a touch screen, relatively larger screen, or a mini-keyboard, it will be regarded as PDA which is outside the scope of this study. Weiss [28] proposed that if a smartphone includes the features of a PDA, it is called a "communicator", which is outside the scope of this study. In our research, we have limited our scope to Java-enabled [11] mobile phones which can execute the Java application developed by J2ME [29]. They do not have powerful processors and only contain limited memory. Also, they are physically small (around 105x45x20mm), and their screens (not touch-sensitive) have limited resolution (around 128x128 pixels).

There are some differences of the hardware among desktop computer, PDA and smartphone. Table 4.1 shows the comparison among these environments, and lets us know the weaknesses of smartphone in terms of hardware. According to this information, the usability problem in smartphone can be investigated.

	Smartphone	***Desktop Computer System***	***PDA***
Input System	12-key keypad with two soft keys and direction key	Mouse and Keyboard	Touch Screen with pen (and mini keyboard in some model)
Output System	Colour/Mono-chrome Screen	Colour Screen	Colour/Mono-chrome Screen
	Very small Screen	Big Screen	Small Screen
	Sound Effect	Sound Effect	Sound Effect
CPU Speed	Slow	Very fast	Fast
Memory Size	Small	Large	Medium
Secondary Storage	N/A	Huge	N/A

Table 4.1 The Hardware in Different Environments [28]

4.1.3 Usability Issues of Smartphone

There are some usability problems for the software application in smartphone, because of limitations in this environment. The developers should be aware of these problems, which were the hypotheses of our experiment presented in next chapter.

4.1.3.1 Long Processing Time

The calculation power of the processor inside a mobile phone is not powerful [25]. Therefore, the software application can only contain simple functionality. Otherwise, if the function is complex, processing will be very slow.

4.1.3.2 Time Delay for Network

As the wireless network of smartphone (GPRS - General Packet Radio Service, one of the GSM data transmission techniques) is relatively slow and unstable [11] compared with the environment of desktop computer, there is a large delay when the application is connected to the server for data exchange. At the same time, because of its limited memory capacity, the developer should reduce the size of the application by increasing more information exchange between server and the application. Thus, the increased network traffic may cause more time delay.

4.1.3.3 Single-handed Operation

In order to have high mobility, the design of the software application should be suitable for single-handed usage. The application is usually used when users are walking or handling other items. As one of his two hands is occupied by maintaining his balance during walking or holding something, such as a bag, the user will find it inconvenient if the application on a smartphone is required to be controlled by both hands.

4.1.3.4 Inefficient Input Options

Smartphone users usually input characters into the software application using keypad. Compared with the desktop computer which has a large keyboard and a mouse, users need to press the same key several times in order to input one character. Thus, the input method in this environment is neither efficient nor user-friendly.

4.1.3.5 Limited Screen Size

Because of the high mobility requirement of smartphone, it needs to be small in size, and therefore usually has a small screen [25]. As a result, users need to scroll the page up and down frequently as the screen cannot display a lot of information. This is inconvenient because the users need to pay close attention in order to read the information from this small screen. This situation is quite dangerous in outdoor environment, as users' eyes only focus on the screen and usually not concentrate on their surroundings. Thus, users can only pay less attention to the application [23].

Sound effects can play a significant role for improving this situation in order to reduce the workload of users' eyes. During long processing of the software, smartphone can generate various sounds so that users need not focus on the screen during waiting time and their eyes can pay attention to the environment.

4.2 Usability Metrics for the Software Application on Smartphone

Many existing usability metrics from previous researches focused mainly on the WIMP software and web application [9, 14]. However, the environment for the software in desktop systems is quite different compared with the operating environment on smartphones. After studying the existing usability metrics [6, 7, 8, 9, 14, 22], some metrics were adopted if they are appropriate for this new environment. Some new usability metrics have also been tailored for the software application in this platform. The details of our proposed metrics are presented in the following section.

4.2.1 Background of Proposed Usability Metrics

Eleven usability metrics for the software application in smartphone environment were proposed. Their background information is described in detail in order to show their importance and the relation among those metrics. This enables better understanding of the usability in this environment. Then we describe which usability attribute should be mapped to each metric.

4.2.1.1 Extra Help (EH)

When using the software application in smartphone, users may face some problems such as not knowing what the next step is or how to use the function. These problems can be reduced by providing a good user interface, which gives an interactive channel to users to be instructed what they should do on the next step. However, if they have no idea or cannot guess their next step, they will seek help from other medium. For example, they may read the user guide of the software application. In addition, users may seek help from the person who have more experience in using that application.

The number of "extra help" was determined by observing how many times user sought help from different medium; except the help function in the software itself because this help function is part of the whole software application design. If this number of "extra help" can be reduced, this implies that users can have a higher chance to complete a task. The design of the software application should have enough instruction or indication for users to complete their task and reach their goals. Therefore, this metric is mapped to the Effectiveness attribute as it measures the accuracy and quality of tasks. The software is highly effective if the number of seeking help is small.

4.2.1.2 Critical Error (CE)

When using software application of smartphone, users may be stopped by the application because of some unexpected events such as defects. If these defects appear in the critical functions in a task, users may not complete it. Then the users may need to repeat the task. We define this type of error as critical error. Critical error is counted if a user took a step backward or exited the current track during usage of the application. They hesitated to continue their procedure owing to that unexpected events [15]. If the number of critical errors can be reduced, this implies that users can have a higher chance to complete their tasks and reach their goals. Therefore, this metric is mapped to the Effectiveness attribute as it measures the accuracy and quality of tasks. The software is highly effective if the number of critical errors is small.

4.2.1.3 Recalled Functions (RF)

Each application comes with multiple functions, but not all of them will be used. If the design of some function is “bad”, users may not be able to identify the functions nor have deep impression about them because users do not know how to use them. Thus, they cannot recall some functions even though the software application has been used a long time. This metric checks the proportion of the functions which can be recalled by users. If the proportion is high, most of the functions can be used and understood by users. The functions in the software application should be designed properly so that users can remember how to use the software easily. Therefore, this metric is mapped to

the Effectiveness attribute, as it measures the accuracy and quality of tasks by checking the level of user understanding of the software application.

4.2.1.4 Key Input Number (KIN)

The keypad is the major input system for smartphone. Inputting characters or strings into the software application via the smartphone keypad is necessary for certain operations. However, this action is very time consuming and troublesome because of the limitation of the keypad. For example, the worst case for inputting one character is to consecutively press the same key four times (e.g. press the "9" key four times to enter the character 'Z'). In addition, if two characters are input sequentially and they are allocated to the same key (e.g. Characters 'T' and 'V' are on the "8" key), the user is required to pause one to two seconds in between inputting these two characters. Therefore, special attention should be paid to this when designing the input environment. For example, the application should not require too many key-pressings, and the input system of the application should be smartly designed [21] in order to reduce the number of key-pressings. If the number of key-pressings per task is very large, users need a lot of time for inputting. Thus, this metric is mapped to the Efficiency attribute as it measures the smoothness and the speed to complete a task. In order to have high efficient software, the number of key-pressings should be as small as possible so that users can finish the task quickly.

4.2.1.5 Total Time (TT)

The completion time of a task is always used to indicate the performance of software application. Total time means how much time a user needs to finish a task. The design of the functions affects the total time per task. If the functions are difficult to understand, the users will require a long time to digest, and they may even perform some wrong steps. These increase the total time to complete the task. The design of the software application should take account of this, and the function should be easy to use and understand so that users can save much time when they interact with the software. It can be easily observed that this metric is mapped to the Efficiency attribute as it measures the smoothness and the speed to complete a task. The software is highly efficient if the total time is small.

4.2.1.6 Response Time (RT)

In smartphone, the performance of the processor is not high compared with desktop computer system. When a user executes a command or a function of a software application in smartphone, that command needs some time to process and the user needs to wait until the process is completed. Response time means the time interval between the starting and the end of the execution of the function. After execution is completed, the user can control the application and give other command to the application again. If the function is very complex, the response time will be quite long. When designing the software application in smartphones, we should be aware of the response time of the function and minimize it as much as possible. If the response time per task is very long, users need to wait for a long while completing the task. Thus, this metric is mapped to

the Efficiency attribute as it measures the smoothness and the speed to complete a task. The software is highly efficient if the response time is small.

4.2.1.7 Frustration Expression (FE)

Users have different feelings towards the application when they are using it. If they can finish a task, or reach some of their goals, they feel satisfied or successful. However, if they cannot achieve their wishes, they will feel annoyed or helpless. They express their feelings by their facial expressions. Frustrated expression is chosen rather than satisfied expression, because users seldom have significantly different expressions when they can reach their goals by completing their tasks as expected. On the other hand, if they cannot finish the task, or do not know how to use the software, their dissatisfaction will be shown by their frustrated facial expression. If the number of frustrated expression is low, this may imply that users feel comfortable when they are using the software. Therefore, this metric is mapped to the Satisfaction attribute as it measures the emotion of the users towards the software application.

4.2.1.8 Single / Two Hand Control (STHC)

In some smartphone software applications, users are required to use both hands to control the application. For example, users may be required to use one thumb to control the direction key and another thumb to press the soft key at the same time. This may cause problems because human often move their hands to maintain balance during walking. Normally, moving hands is not too critical to maintain balance during walking when one is paying attention during walking. If they are required to use both hands for

the key input, it will be difficult to control their body movement. The situation should be better if users can have one free hand for balancing during walking. Therefore, this metric is mapped to the Satisfaction attribute as it predicts the emotion of the users towards the software application. In order to develop highly satisfactory software, the software application should be designed to be operated by single hand.

4.2.1.9 Software Size (SS)

As the memory for software application is quite limited on smartphone, the size of the software application becomes very significant. If it is too large, most of the memory space in smartphone will be occupied so that users cannot import other useful applications into the smartphone. It will be more convenient if the application size is small as users need not pay much attention on the spacing issue. In addition, the software size is a good indicator of the complexity of the software. If its size is large, the application may be too complex and then this application may execute slowly as the processor in smartphone is not powerful. Therefore, this metric is mapped to the Satisfaction attribute as it predicts the emotion of the users towards the software application. In order to have high satisfaction software, the size of the software application should be as small as possible.

4.2.1.10 Page Colour (PC)

Most smartphones have colour screens nowadays. Software developers can design better layout on colour screen phones rather than on mono-colour screen phones. Developers should be aware of the number of different colours per page on the colour

screen. If this number is not enough, users will be bored and think that the screen layout is not very clear. On the other hand, if there are too many colours, users will think that the layout is too messy. As a result, the number of colours used per page can be used in order to evaluate the usability of the software. The standard number of colours used per page can be based on some design guidelines [14]. This metric, Page Colour, is a good indicator for users' feelings towards the software application and should be mapped to the Satisfaction attribute. The number of colours for screen layout should be near the reference value of page colour so that users feel comfortable when they are looking at phone screens.

4.2.1.11 Sound Effect (SE)

Smartphone does not have many output channels. As a result, sound effect provides a very significant interaction between software and users in this environment. If the software application provides appropriate sound effects, the usability level of the software can increase. For example, when the application executes a long computation, users need to wait. This is not convenient. If they paid attention to other tasks when waiting, such as walking, they cannot deal with the completion of the execution at once. Sound effects can be a notification to let users know the process had ended. Thus, this metric should be mapped to the Satisfaction attribute since sound effect can provide additional convenience to users. If most of the functions in software application support sound effect, users can control it easily as they have this extra channel to interact with the application.

4.2.2 Computation of Proposed Usability Metrics

The computation of the proposed metrics is presented in this section. The basis for computation is the five metric conversions discussed in section 3.4. The parameters and inputs of each of these metrics will be described in detail and then an example is shown to illustrate the computation.

4.2.2.1 Extra Help (EH)

This metric uses Deduce Base Conversion for its calculation. The number of users seeking extra help per task is "IV", which can be investigated by counting how many times a user seeks help from other medium per task during a usability evaluation. "α" can be obtained via the raw data extracted from questionnaires or testing results of previous mobile device research. One of the examples of the raw data is the maximum times for a user seeking help when they cannot continue to execute a task. This data can then be used to compute the value of α. Let IV be 8 and α be 8. In this case, the upper case of the formula in the conversion is used for calculation since $\alpha \times IV < 100$. The output value of this metric is 36.

4.2.2.2 Critical Error (CE)

This metric uses Deduce Base Conversion for its calculation. The number of critical errors per task is "IV", which can be investigated by counting how many critical errors generated per user per task during a usability evaluation. "α" can be obtained via raw data from questionnaires or previous testing results extracted from other mobile devices, such as the maximum number of critical errors that will cause users to reject a task. This

data can then be utilized to compute the value of α. Let IV be 6 and α be 5. In this case, the upper case of the formula will be used for calculation since α x IV < 100. The output value of this metric is then 70.

4.2.2.3 Recalled Functions (RF)

This metric uses Portion Conversion for its calculation. The number of recalled functions in the software application is "IV", which can be determined by asking a user how many functions he/she can remember after the usability test. The total number of functions in the software application is "TIV", which can be obtained by counting the number of functions in the specification of the application. Let IV be 7 functions and TIV be 10 functions. The output value of this metric is then 70.

4.2.2.4 Key Input Number (KIN)

This metric uses Standard Dependent Conversion with SMAX for its calculation. The number of key-pressings per task is "IV", which can be determined by counting the number of key-pressings by a user per task during a usability evaluation. The standard value of number of key-pressing is "SMAX", and "UB" is the maximum number of key-pressings which is acceptable by a user. These can be obtained by previous testing result or raw data from questionnaires. Let SMAX be 5 keys, UB be 45 keys and IV be 25 keys. In this case, the middle case of the formula will be used since SMAX ≤ IV < UB. The output value of this metric is 50.

4.2.2.5 Total Time (TT)

This metric uses Standard Dependent Conversion with SMAX for its calculation. The total time per task is "IV" which can be obtained by noting how much time a user took to complete a task. We started to tally IV when the user had just begun to go through a task and stopped it when all objectives of the task have been completed. The standard value for total time per task is "SMAX", and "UB" is the largest completion time which is acceptable by a user. These can be obtained by previous testing result or raw data from questionnaires. Let SMAX be 15 seconds, UB be 45 seconds and IV be 55 seconds. In this case, the lower case of the formula will be used since IV > UB. The output value of this metric is 0.

4.2.2.6 Response Time (RT)

This metric uses Standard Dependent Conversion with SMAX for its calculation. The response time per task is "IV", which can be obtained by observing how much time a user waited for the application when the function is in process. All the waiting time within one task will be summed up to give IV. The standard value of response time per task is "SMAX", and "UB" is the longest waiting time which is acceptable by a user. These can be obtained by previous testing result or raw data from questionnaires. Let SMAX be 10 seconds, UB be 30 seconds and IV be 8 seconds. In this case, the upper case of the formula will be used since IV ≤ SMAX. The output value of this metric is 100.

4.2.2.7 Frustration Expression (FE)

This metric uses Deduce Base Conversion for its calculation. The number of frustrated facial expression per task is "IV", which can be determined by counting how many times users have frustrated expressions per task during a usability evaluation. Video captures can be used so that the number of frustrated expressions can be counted accurately during playback. "α" can be supplied via raw data from questionnaires or previous testing results extracted from other mobile devices, such as the maximum number of unsatisfied feelings when users cannot accept a task. This data can then be utilized to compute the value of α. Let IV be 10 and α be 4. In this case, the upper case of the formula will be used for calculation since $\alpha \times IV < 100$. The output value of FE is 60.

4.2.2.8 Single / Two Hand Control (STHC)

This metric uses Portion Conversion for its calculation. The total time when a user is using the software application single-handedly is "IV", which can be determined from counting how much time users are using the software with one hand per task. The total time for using the software application is "TIV" which can be obtained by timing the total time users are using the software application per task. Let IV be 60 seconds and TIV be 80 seconds. The output value of this metric is then 75.

4.2.2.9 Software Size (SS)

This metric uses Standard Dependent Conversion with SMAX for its calculation. The size of the software application is "IV", which can be determined from the size of the software code. The standard size of the software application is "SMAX" and "UB" is the largest size which is acceptable by a user. They can be supplied by previous testing

result or raw data from questionnaires. According to these data, some useful information can be abstracted such as the number of applications which users are preferable, and the minimum number of applications which a user usually imports into a smartphone. These numbers, combined with the standard memory size in smartphone, can supply the calculation of SMAX and UB. Let SMAX be 100kbytes, UB be 300kbytes and IV be 160kbytes. In this case, the middle case of the formula will be used since SMAX ≤ IV < UB. The output value of this metric is then 70.

4.2.2.10 Page Colour (PC)

This metric uses Reference Dependent Conversion for its calculation. The number of colours on the screen layout is "IV", which can be obtained by counting how many different colours are on every screen layout in the task. Then the average of these numbers of colour on different layouts is calculated to be IV. The reference value of the number of colours used on the screen layout is "RV", and the maximum and the minimum of colours per screen are "UB" and "LB" respectively. They can be obtained by raw data from surveys or statistics. Let RV be 5 colours, UB be 10 colours, LB be 2 colours and IV be 6 colours. In this case, the lower case of the formula will be used for calculation since RV < IV ≤ UB. The output value of this metric is 80.

4.2.2.11 Sound Effect (SE)

This metric will use Portion Conversion for its calculation. The number of functions which had included sound effect is "IV", which can be determined by counting how many functions support sound effect. The number of functions which could use sound

effects in the software application is “TIV”, which can be obtained by estimating the number of functions in the software application should be implemented sound effect. The specification of the software application can be evaluated and this also can help to find TIV. Let IV be 3 places and TIV be 6 places. The output value of this metric is 50.

4.3 Summary

Table 4.2 gives the summary of the usability metrics used in this study. Column 1 provides the name of metrics and their brief description is provided in Column 2. Column 3 lists the related usability issues of the metrics. The usability issues discussion in section 4.1.3 are very significant in smartphone environment. In fact, some of the other common traditional usability issues, such as ‘Design of Layout’ and ‘Design of Procedure’, were not included in that section. The previously mentioned usability issues are shown in italic font. Finally, the usability attribute, to which the metric belongs, is provided in Column 4.

Metric Name	***Metric Description***	***Usability Issue***	***Attribute***
Extra Help (EH)	The times of seeking help per task from the medium other than in the application	Design of Layout, Design of Procedure, Design of Online Help	Effectiveness
Critical Error (CE)	Number of critical errors per task	Design of Layout, Design of Procedure	Effectiveness
Recalled Function (RF)	Number of functions which can be recalled by a user	Design of Layout, Design of Procedure, Design of Application Functionality	Effectiveness
Key Input Number (KIN)	Number of key pressing per task	*Inefficient Input Options*	Efficiency
Total Time (TT)	Total completion time per task	*Long Processing Time*, *Time Delay of Network*, Design of Layout, Design of Procedure	Efficiency
Response Time (RT)	The time for waiting the process per task	*Long Processing Time*, *Time Delay of Network*	Efficiency
Frustration Expression (FE)	Number of frustrated facial expression from user per task	Design of Application Functionality	Satisfaction
Single / Two Hand Control (STHC)	The software application mainly controlled by one hand or two hands	*Single-handed Operation*, *Limited Screen Size*	Satisfaction
Software Size (SS)	The size of the software application	*Long Processing Time*	Satisfaction
Page Colour (PC)	The average number of colours of screen layout	Design of Layout	Satisfaction
Sound Effect (SE)	The software application mainly supported by sound effect or not	*Limited Screen Size*	Satisfaction

Table 4.2 The Usability Metric List for Smartphone

Our proposed usability metrics in smartphone environment are designed for different *levels* of measurement. There are three levels of measurement: Application, Task and Function. Application level is the highest level of measurement and the metric belonging to this level measures the performance of the whole application; Task level is the middle level of measurement and the metric belonging to this level measures the performance for each task; Function level is the lowest level of measurement and the metric belonging to this level measures the performance each function. Fig 4.2 shows the relationship among application, tasks and functions. One application can have several tasks and each task is composed of several functions.

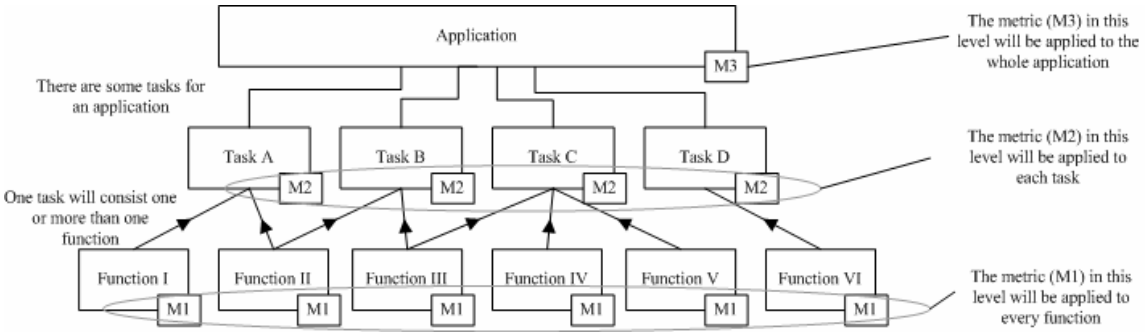


Fig. 4.2 The Metrics Applied in Different Levels of Measurement

Table 4.3 depicts a summary of our proposal usability metrics in smartphone environment, their conversion types and their levels of measurement in the software. Name and abbreviations of each metric is shown in Column 1. The conversion types of each metric (see section 3.4) are provided in Column 2, and Column 3 shows the levels of measurement.

Metric Name	***Conversion Type***	***Level of Measurement***
Extra Help (EH)	Deduce Base Conversion (DBC)	Task Level
Critical Error (CE)	Deduce Base Conversion (DBC)	Task Level
Recalled Function (RF)	Portion Conversion (PC)	Application Level
Key Input Number (KIN)	Standard Dependent Conversion (SDC)	Task Level
Total Time (TT)	Standard Dependent Conversion (SDC)	Task Level
Response Time (RT)	Standard Dependent Conversion (SDC)	Function Level
Frustration Expression (FE)	Deduce Base Conversion (DBC)	Task Level
One / Two Hand Control (OTHC)	Portion Conversion (PC)	Application Level
Software Size (SS)	Standard Dependent Conversion (SDC)	Application Level
Page Colour (PC)	Reference Dependent Conversion (RDC)	Function Level
Sound Effect (SE)	Portion Conversion (PC)	Application Level

Table 4.3 The summary of Conversion Type and Level of Measurement of Proposed Metrics

Chapter 5 Research Questions

In this chapter, the hypotheses of our research are discussed to show our focus in this study. The hypotheses were then verified and validated by aligning the criteria to a specially designed experiment. We classified the hypotheses into two types: General Hypotheses and Hypotheses of Software Comparison. General Hypotheses were proposed according to users' background towards their performance in the experiment. Hypotheses for Software Comparison were proposed according to the characteristics of each tested software which influences users' performance in the experiment. To check for mistakes in our hypotheses and how many proposed metrics are practical to apply with our limited research resources, we first conducted a pilot test before the experiment.

5.1 General Hypotheses

There are many different reasons for users to consider whether or not a software application is highly usable. Some of the reasons are related to users' background or history towards a particular kind of software application. Also, the software application in smartphone environment is quite different from that in desktop computer. According to the constraints and limitations in this environment, we tried to predict the feelings of users. The following General Hypotheses are proposed.

GH1, The user has better performance in the experiment if he/she has used mobile phone for a long period.

If a person has used mobile phone for long time, he/she is quite familiar with the basic functions in mobile phone, such as phonebook or organizer, during mobile situation such as walking. This experience benefits him/her to grasp the software usage in smartphone quickly so that his/her performance may be better in the experiment.

GH2, The user has better performance in the experiment if he/she has prior experience of using WAP(Wireless Application Protocol) phone or smartphone.

If a person had previously used WAP phone to access the Internet, he/she will have experience to use the mobile phone internet browser which has similar controls as the software application in smartphone. 'Soft Key' is the main input medium for both of them. If a user already has experience of using smartphone, he/she can finish the tasks very smoothly compared with other users who have not, as the environment is not new for him/her.

GH3, The user has better performance in the software of Nokia smartphone if he/she has experience using Nokia mobile phone before.

In our experiment, there were two software applications in smartphones for user to test. (The detail description of these two software applications will be presented in next section.) One of them was installed in a Nokia smartphone and another was installed in a Siemens smartphone. If a person had used Nokia mobile phone before, he/she can be familiar with the control interface of Nokia smartphone because screen layouts and control flows between Nokia mobile phone and Nokia smartphone are quite similar.

GH4, The user has better performance when using the next software application after using the previous one.

If a person has never used smartphone before, he/she need some time to warm-up in order to become familiarized with this new platform. When using the software application for the first time, the user is still in the learning process, and therefore he/she may make a number of mistakes or uses more time to complete the tasks. As a result, his/her performance of next software application has improvement compared with the previous one.

GH5, The user has better performance for the task which requires user inputting characters if he/she usually sends SMS.

Inputting characters is usually regarded as one of the least user-friendly tasks in mobile device because of the limitation of the input interface. User usually needs to press the same key several times in order to input one character. However, if a user always sends SMS, he/she has already been well trained for inputting characters in this environment. Therefore, he/she has better performance for the task which requires him/her to input characters.

GH6, The user has better performance in the experiment if he/she has IT background.

If a person has studied Computer Science, Information Technology related degree or certificate before, or is working in Information Technology firm with related position, he/she may have much experience in using different kinds of software application. He/she may also understand the concept or the logic of how software works. Although the software application in smartphone environment may be quite new for him/her, his/her previous knowledge let him/her pick up the application quickly.

5.2 Hypotheses for Software Comparison

In this research, we conducted experiments to test two different software applications, with each application having its own characteristics which may influence its usability. Therefore, the characteristics and assumptions of each application are described first. Then the Hypotheses for Software Comparison are proposed. We also predicted which software application may have better usability result.

5.2.1 Assumptions of the Software Applications

Two different software applications were available for usability tests: *M-Travel* and *Movie on Phone*. Fig. 5.1 shows the screen layouts of these two applications. *M-Travel* has a lot of functions related to travel agent and contains a lot of tour information. Users can login to the system to book their preferred tours. *Movie on Phone* has information about which cinemas are showing which movies. After a user login to the system, he/she can buy movie tickets by the smartphone. Before conducting the experiment, we did some comparisons between these two applications.

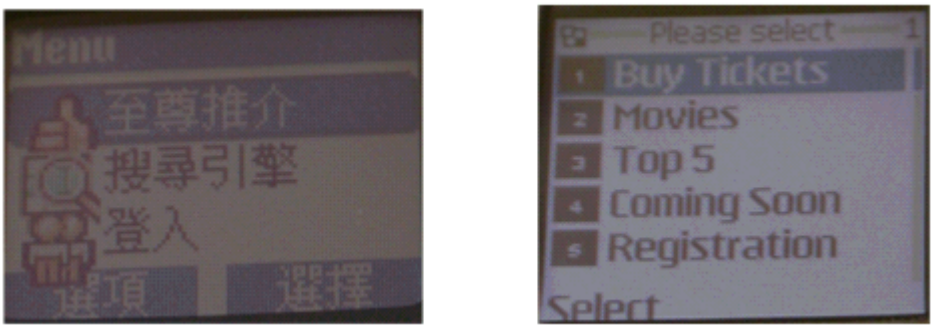


Fig. 5.1 The Screen on the Left is M-Travel, and the Screen on the Right is Movie on Phone

First of all, for the interface, M-Travel is better because its layout has colourful icons so that users can recognize the functions easily, and some graphical items which can be quite attractive for usage. In addition, M-Travel has a Chinese language interface which favours users with Chinese as their first language. Furthermore, only M-Travel supports sound effect during usage, thus allowing user interaction by another output channel.

One of the characteristics of the software application in smartphone is the significant time delay compared with that in desktop computer. Both of the applications required time to obtain information though network. Movie on Phone has a longer network delay and does not give feedback during this waiting period. Thus, this may cause disadvantages for usage.

Finally, the control flow of Movie on Phone is a ‘step-by-step’ approach which is quite simple. Usually, there are only up to two functions per softkey, and users only follow the control flow step-by-step to complete the task. For example, user must input some data first, and then some items are output for selection. The control flow of M-Travel is an ‘all-in-one’ approach, which means each softkey has a designated function menu which advantages fast selection. User has greater feasibility to control the application in

order to acquire what he/she needs and can complete the task faster as there is no need to go through unnecessary steps.

5.2.2 Hypotheses and Predicted Results

SH1, The colourful screen layout with graphics is more attractive than plain text layout.

User usually feels bored when he/she faced a plain text layout which is not attractive. If the screen layout has some colourful graphics, this can raise user's interest to use the application. In addition, user can recognize the functions more easily in the graphical layout compared with that in plain text layout. Therefore, M-Travel which has colourful graphics layout should be more usable than Movie on Phone in this area.

SH2, The processing time and response time of the application is significant for the usability of the software.

If users need to wait a long time for a software application to execute, this diminishes the efficiency to complete the task. Also, users do not feel happy regarding the delay because it highly affects the smoothness to finish the task. The processing time and response time of Movie on Phone are quite long and it does not provide any feedback during the waiting period. Although M-Travel also has some processing time and response time, it is faster than that of Movie on Phone. In addition, during the waiting period, M-Travel provides messages to inform users waiting for the processing. Therefore, M-Travel should be more user-friendly, as the waiting time is a bit shorter and users can expect to wait for a while because of the notification.

SH3, Users feel difficult to walk when they are using the application in smartphone concurrently.

It is common for a person using the software application in smartphone when they are walking. However, when he/she is using the application, he/she pays less attention for walking. As a result, the user feels difficult controlling his/her movement and balancing his/her body during walking. A user may pay less effort to use M-Travel as he/she can recognize the functions easily because of its graphical layout. Therefore, M-Travel should be better than Movie on Phone in this usability aspect.

SH4, Users find it convenient if they can use only one hand to control the application in smartphone.

When a user is using the software application in a smartphone, he/she is always in the highly mobile situation, such as walking on the street. The user seldom has two idle hands to control the software application in smartphone. For example, one of his/her hand needs to take a bag, or keeps his/her body balance. Also, if both of his/her hands are used to control the application, it is difficult to balance his/her body during walking. Therefore, the application should be designed to be used by one hand only. Movie on Phone should be better in this usability aspect as the use of softkey is more consistent than that of M-Travel.

SH5, Sound effect plays a significant role in the smartphone.

Because of the limitation of the output display of the software application in smartphone, other output medium should be introduced to this kind of application in order to enhance the interaction between human and the application. Sound effect is a common and simple way to notify a user about the current state or the response of this application. As a result, the application which supports sound effect, such as M-Travel, can enhance the usability of the software.

SH6, If users require a lot of extra help, the usability of the software application is not satisfactory.

A user asks for extra help if he/she faces a problem which cannot be solved by himself/herself. He/she may get lost or does not know what the next step is. This problem is usually caused by the poor design of the application and then user was misled, and cannot predict the flow of the software. Although the user can continue his/her task after obtaining this extra help, the smoothness of task completion can be affected if he/she requires a lot of extra help.

SH7, The usability of the software application is not satisfactory if the number of key-pressing per task is too many.

In smartphone platform, keypad and softkeys are the main input interface to control the software application, and this environment does not have touch screen or mouse-like input device. Therefore, the number of key-pressings is a significant usability factor in this platform. If the design of the application is user-friendly, the unnecessary key-pressings should be minimized, and the user only needs to press a few keys to finish a task.

SH8, Users only show their frustrated expressions which represents their dissatisfaction.

Users show their personal feelings on their faces sometimes. However, they seldom have any facial expressions even though they are satisfied when they finished a task using the software application. On the other hand, if there are some difficulties during usage, their faces may show frustration, because they usually expect they can finish a task smoothly. As a result, their frustrated expressions can be used to show the usability level of the software.

SH9, 'All-in-One' approach is better than 'Step-by-Step'.

'All-in-One' approach is to put all necessary functions in the softkey menu in order to let users use the function they want by pressing only few keys. 'Step-by-Step' approach is to minimize the number of functions in the softkey menu, and the user should follow many simple steps. We assumed that 'All-in-One' approach is better, as the control feasibility in the application is better and users find it more convenient. Therefore, M-Travel which uses 'All-in-One' approach is more usable than Movie on Phone which uses 'Step-by-Step'.

5.3 Adjustment after Pilot Test

Before we conducted the full experiment to verify our proposed metrics and hypotheses, a small scale pilot test was conducted in order to fine-tune our proposed hypotheses. We would like to ensure that they are correct and make sense before moving onto the full experiment. At the same time, we tried to estimate how much metric data that we will be able to distill with our limited resources. Five subjects have participated in the pilot test and they did the similar procedures in the normal test (Please refer to Chapter 6) .

However, during this pilot test, each subject was asked their feelings and comments frequently (this would not happen in the normal test) in order to investigate any potential problem in the test. Thus, the test would be pause if we need to discuss the improvement of the test. Finally, we modified some wording in the scenario sheet and simplified some of the tasks. Each subject spent around 1.5 hours in the pilot test.

After the pilot test, we found that most of our proposed hypotheses did not have obvious problems. Although our subjects did not strongly agree to all of our hypotheses as different people have different concerns or criteria, they also find that most of our hypotheses were reasonable except SH9. Most of our subjects in the pilot test thought that the design of control in Movie on Phone, which was 'step-by-step' approach, is less complex compared with M-Travel. As a result, we modified SH9 and Table 5.1 shows all of the hypotheses we used in the full experiment.

Abb.	*Hypothesis*
GH1	The user has better performance in the experiment if he/she has used mobile phone for a long period.
GH2	The user has better performance in the experiment if he/she has prior experience of using WAP (Wireless Application Protocol) phone or smartphone.
GH3	The user has better performance in the software of Nokia smartphone if he/she has experience using Nokia mobile phone before.
GH4	The user has better performance when using the next software application after using the previous one.
GH5	The user has better performance for the task which requires user inputting characters if he/she usually sends SMS.
GH6	The user has better performance in the experiment if he/she has IT background.
SH1	The colourful screen layout with graphics is more attractive than plain text layout.
SH2	The processing time and response time of the application is significant for the usability of the software.
SH3	Users feel difficult to walk when they are using the application in smartphone concurrently.
SH4	Users find it more convenient if they can use only one hand to control the application in smartphone.
SH5	Sound effect plays a significant role in smartphone.
SH6	If users require a lot of extra help, the usability of the software application is not satisfactory.
SH7	The usability of the software application is not satisfactory if the number of key-pressing per task is too much.
SH8	Users only show their frustrated expressions which represents their dissatisfactory.
SH9	'Step-by-Step' approach is better than 'All-in-One'.

Table 5.1 The hypothesis list for the full experiment

As our resource was limited, we could not extract too many different kinds of data for our proposed metrics during the experiment. At the same time, we only tested two software applications; therefore, the comparison of metric results among software applications was not enough for the metrics which belong to Application Level or Function Level (refer to chapter 4.3). As a result, we only chose the metrics which belong to the Task Level (refer to chapter 4.3) because each software application

contained several different tasks for us to make the metrics comparison. Table 5.2 shows all of the metrics we used in the full experiment.

Abb.	***Metric***
EH	Extra Help
CE	Critical Error
KIN	Key Input Number
TT	Total Time
FE	Frustration Expression

Table 5.2 The metric list for the full experiment

Chapter 6 Experiment Design

The objective of this experiment is to collect quantitative data from users' behaviour when they are interacting with the software applications. Then these data can determine the usability level of each software application via our proposed usability metrics. At the same time, the comments from subjects had been consolidated in order to analyze which usability problem or weakness is the most serious from their point of view. This information can be a good reference for developers to design and develop software in this platform. Finally, the result of our usability metrics was compared with users' comments from questionnaires in order to validate our metrics.

6.1 Background of Subjects

Sixty subjects participated in this experiment and 30 of them were male, and 30 were female. Their ages range from 18 to 37. All of them are current mobile phone users and they had used mobile phones for around 4 years on average. Around 57% subjects have information technology education background, and those people are the potential users of this kind of application. 48% of our subjects had used Nokia mobile phones before, and they might have bias in our experiment as one of the applications was installed on Nokia smartphone.

6.2 Experiment Procedure

Our subjects were required to use two different software applications: “M-Travel” which ran on Siemens S57 and “Movie on Phone” which ran in Nokia 7250 (see Appendix A). The evaluation sequence for these applications is random. Therefore, half of our subjects evaluated M-Travel first, while the remaining subjects used Movie on Phone at the beginning in order to minimize the bias according to hypothesis GH4 (section 5.1). Only one subject is allowed to conduct this experiment at each time slot, and he/she required around forty-five minutes to complete it on average. There are many steps in our experiment which included: briefing, scenarios for the application usage, question and answer (Q&A) session for each scenario, questionnaires of users’ comment, and personal information of each subject. Fig. 6.1 provides a full picture of the experiment. Each application has 3 scenarios, and each scenario has a Q&A session afterwards. Users’ feelings and comments could be obtained immediately via this interactive arrangement.

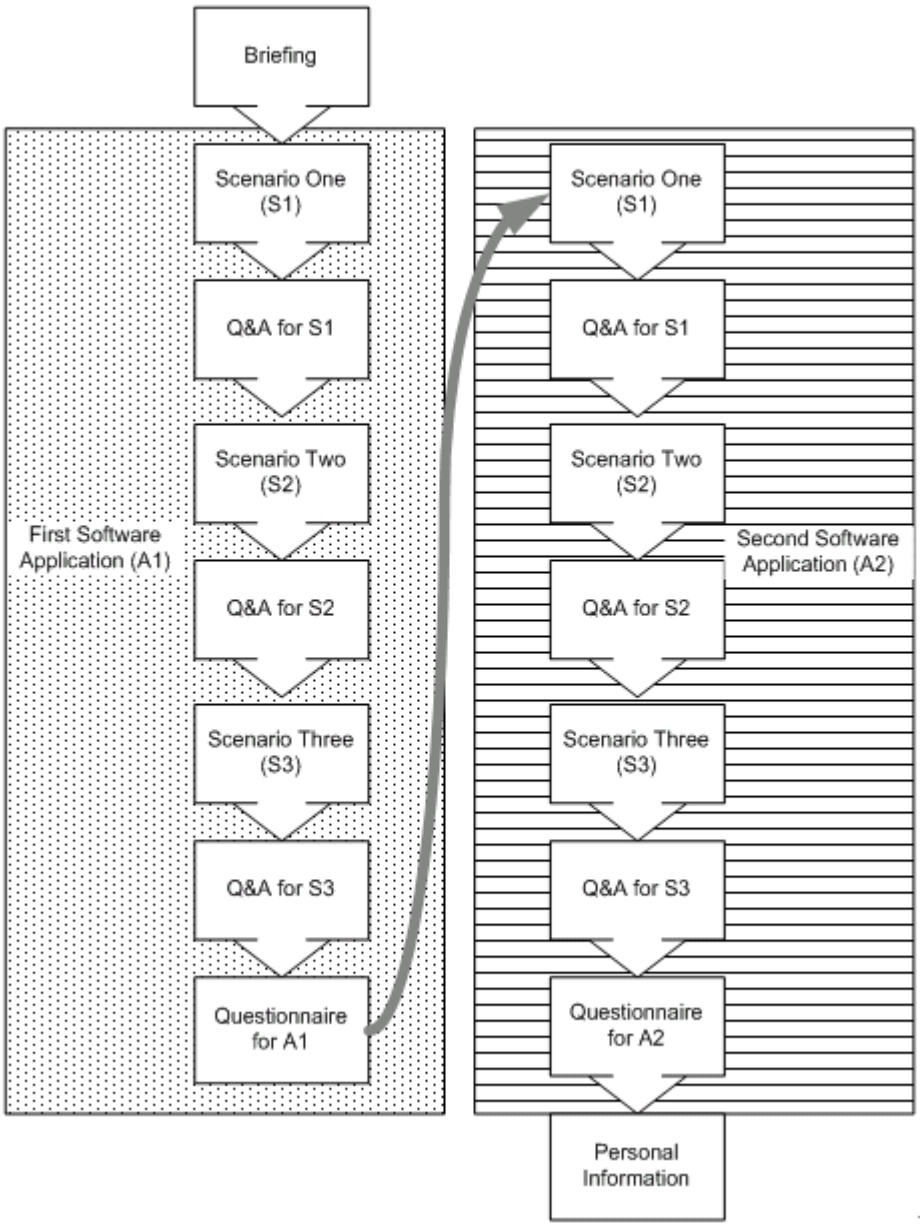


Fig. 6.1 The Procedure Chart of the Whole Experiment

6.2.1 Briefing Session

A briefing would be given at the beginning of the experiment. This included: the experiment objective, its schedule, the concept of usability, reminders during the experiment and a simple reference guide on how to control the software. Some card

boards, which contained key points of the briefing session, were placed on the desk in order to remind a subject when he/she was using the software application.

6.2.2 Scenarios for Application Usage

There were two software applications in our experiment, and each of them was tested by a particular set of scenarios. Both sets had the same structure which contained three scenarios. In order to obtain accurate and detail feedback, all subjects were required to be 'thinking aloud' [28] during this session so that they would tell us what was going through in their minds when using the application.

The first scenario allowed users to become familiar with the application. They were only required to browse the application and find out what functions it contained. The subjects were also requested to walk along the corridor (Fig. 6.2) when they were handling this scenario, because this simulated a high mobility environment which gave them a feeling of real situation. An observer followed the subjects and used digital video camera to capture their behaviours and comments.

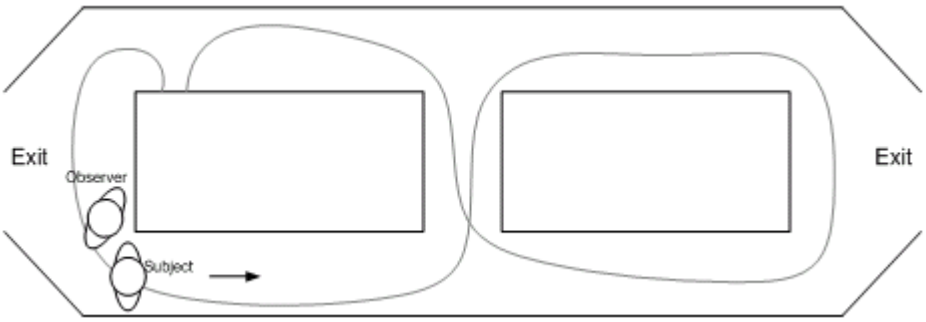


Fig. 6.2. The Walking Path had Many Other Persons Pass by and the Subject was Requested to Make Few Turns.

In the next scenario, the subjects sat in front of a table as depicted in Fig 6.3. The objective of this scenario was to login to the system. This action is an essential step for

e-commerce application. The subjects needed to correctly input the user login and password in order to complete this scenario.

In the final scenario, the subjects should search and locate the appropriate product, and try to order it. Conducting transaction in the application is also an essential part for e-commerce application.

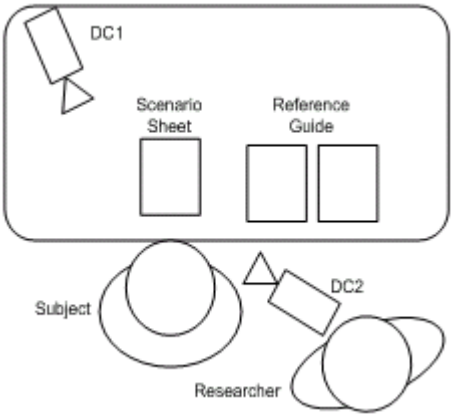


Fig. 6.3 The Layout of the Setup for Scenario 2 and 3

Most of the quantitative data were obtained during this session, such as completion time and number of key-pressings. Fig. 6.3 shows the equipment setup in this session. The subject was captured by two digital video cameras: DC1 captured facial expressions and recorded his/her comments during the whole session; DC2 captured his/her finger-movements and number of key-pressings when he/she was performing the tasks. The scenario sheet (see Appendix C) was placed in front of the subject so that he/she knew what the next step was when using the application. A researcher observed how the subject controlled the software application and provided hints if he/she got loss. In order to minimize the disturbance towards the subject, the researcher stood behind him/her.

6.2.3 Question and Answer Session

After a subject completed each scenario, he/she was asked about his/her feelings or comments towards the tasks or functions in that scenario. Sometimes subjects might raise useful suggestions to improve the usability of the application. Since they only handled a few tasks before each Q&A session, their memories were still very fresh regarding their feelings for those tasks. The arrangement for Q&A session could ensure that we obtained all the useful information from users' feelings and comments before they had forgotten their points.

6.2.4 Questionnaires and Personal Information

After a subject had completed all the scenarios and Q&A sessions of each application, he/she should fill in a questionnaire (see Appendix B) to rate the usability level of that application. This questionnaire was designed to extract users' comments on a five-point scale. The result of the questionnaire was then compared with the result of our metrics in order to validate our proposed usability metrics.

At the end of the experiment, the subject then needed to fill in another questionnaire (see Appendix B) which is about their background information. This personal information was used to analyze the validities of our general hypotheses.

Chapter 7 Experimental Result and Finding

There were two parts in our usability testing: Quantitative Analysis and Qualitative Analysis. Quantitative experimental findings were about the result of the proposed usability metrics in this study. The users' comments and suggestions about the usability of the software application were classified into qualitative result.

7.1 Quantitative Analysis

The objective of this analysis is to justify the validity our proposed usability metrics. At the same time, the result from these metrics was compared with the users' background information in order to verify our hypothesis. The statistical software Application, SPSS, was used to analyze the quantitative data. The statistical methods, such as *t*-Test, Correlation and Chi-Square (x^2) were applied in the analysis. The resultant data from the SPSS are included in Appendix D in this research.

7.1.1 Comparison of M-Travel & Movie on Phone

There were two scenarios for each application and we summed up all the marks from four usability metrics (NSH, TT, FE and TE). Therefore, the overall total for each application should be 800 (100 x 4 x 2). Then we used *t*-Test to compare the mean of the overall marks from the two software applications. The paired-samples *t* test analysis indicates that for the 59 subjects, the mean of total score of the "M-Travel" (M = 396.21) was significantly lower at the $p < .01$ level ($p = .004$) than the mean score of the "Movie on Phone" (M = 427.28). As a result, the usability level of "Move on Phone" is better and this conclusion is the same as the users' preference during the interview. The

statistical findings also show that a significant correlation exists between these two variables ($r = .398$, $p < .01$). This also suggests that hypothesis SH9 ('Step-by-Step' approach is better than 'All-in-One') is correct.

7.1.2 Findings of Usability Metrics

The usability metric results from different subjects were analyzed in this section. These results were compared with users' preference during the interview and also other quantitative data from the questionnaires. We used these comparisons to investigate whether or not those usability metrics have relationship with the usability opinion from subjects. As a result, the validity of each metric could be checked by different statistical methods.

7.1.2.1Number of Seeking Help (NSH)

This usability metric was applied in scenario 2 and 3 for each software application. In each scenario, the metric results from each subject obtaining from two different applications were compared in order to find out which application the subject may feel more usable. From this metric result comparison, we classified three kinds of user preference: "M-Travel" (Metric mark higher in M-Travel), "Movie on Phone" (Metric mark higher in Movie on Phone) and "Same" (both marks are similar). After subjects had finished all scenarios, they were required to rank the best software application for each scenario. In each scenario, they also have three choices to choose: "M-Travel", "Movie on Phone" and "Same".

Then the metric result comparison was analyzed with the user ranking which were obtained during the interview. We applied Chi-Square test to those data in scenario 2 and the result was significant [$F(1/56) = 12.082$, $p < 0.05$]. At the same time, the data in scenario 3 was significant, too [$F(1/58) = 11.970$, $p < 0.05$]. Therefore, it was verified and proved that this metric could represent the user feeling towards the usability of the software and the hypothesis SH6 (If users require a lot of extra help, the usability of the software application is not satisfactory) holds.

7.1.2.2 Total Time (TT)

The analysis and the comparison of the result of this usability metric were similar with that of "Number of Seeking Help". The result of Chi-Square test on this metric attained a marginal significant in scenario 2. [$F(1/57) = 8.805$, $p < 0.10$].

At the same time, Regression test was also applied in order to find out whether or not this metric has correlation with the user feeling in the questionnaire. The result of this metric was analyzed with the rankings of different usability items in the questionnaire. A significant correlation existed between this usability metric and rating of the questionnaire item, 'The speed of using it' in the "Movie on Phone" [$r = .344$, $p < .01$]. Another significant correlation existed between the metric and the rating of 'The inputting information time' for "Movie on Phone" [$r = .338$, $p < .01$]. Finally, 'Your overall usability rating' in "Movie on Phone" questionnaire also had a significant correlation with this metric [$r = .277$, $p < .05$]. Therefore, this metric was verified which can represent the user feeling towards the usability of the software and the hypothesis

SH2 (The processing time and response time of the application is significant for the usability of the software) holds.

7.1.2.3 Frustration Expression (FE)

The Regression test was applied to find out whether or not there was any correlation between this metric and subject feeling according to the questionnaire. Similar analysis as the "Total Time" was performed. Significant correlations existed between this usability metric and the rating of 'The design of screen layout' [M-Travel: $r = .285$, $p < .05$; Movie on Phone: $r = .275$, $p < .05$], 'The ease of input control' [M-Travel: $r = .276$, $p < .05$; Movie on Phone: $r = .258$, $p < .05$], and 'Your feeling when you are using it' [M-Travel: $r = .359$, $p < .01$; Movie on Phone: $r = .323$, $p < .01$] in both questionnaires. The analysis result between this usability metric and questionnaire rating of 'Your overall usability rating' in the questionnaire attained the marginal significant [$r = .204$, $p < .10$] for "M-travel" and significant correlation. [$r = .299$, $p < .05$] for "Movie on Phone". Thus, this usability metric was verified which can represent the user feeling towards the usability of the software and the hypothesis SH8 (Users only show their frustrated expressions which represents their dissatisfactory) holds.

7.1.2.4 Total Error (TE)

In previous sub-section (7.1.2.1, 7.1.2.2), we have introduced two of our statistical analysis methods in this study. We could not find any significant relationship or correlation for the result of this usability metric with the other variable. One reason may be there were only two software applications in this test and their usability level in this

area was quite similar. In the future, the number of testing applications should be increased to provide a more complete analysis.

7.1.3 Findings for Hypotheses

Some of our proposed hypotheses could be analyzed by quantitative statistical method. We applied *t*-Test analysis for the following hypotheses. We divided the subjects into two groups and the performance of each subject was reckoned by the overall mark of those usability metrics (please refer to 7.1.1). Then the mean marks of two groups were compared in order to find out the validity of each hypothesis.

7.1.3.1 GH1: The user has better performance in the experiment if he/she has used mobile phone for a long period

According to the statistical analysis, no significant difference between the mean marks of two subject groups exists. Therefore, we could not verify this hypothesis.

7.1.3.2 GH2: The user has better performance in the experiment if he/she has prior experience of using WAP (Wireless Application Protocol) phone or smartphone

In the scenarios of "M-Travel", the independent-samples *t* test analysis indicated that 17 subjects, who did not have any experience on WAP phone, had a mean of 376.64, and 43 subjects, who had experience on WAP phone, had a mean of 403.41, and the means did not differ significantly at the $p < .05$ level ($p = .114$). Levene's test for Equity of Variances did show significant differences ($p < .01$). In the scenarios of "Movie on

Phone", the independent-samples *t* test analysis indicated that 16 subjects with no WAP phone experience had a mean of 407.76 and 43 subjects who had WAP phone experience had a mean of 434.54 and the means did not differ significantly at the $p < .05$ level ($p = .119$). Levene's test for Equity of Variances did show significant differences ($p < .05$). According to the above result, there was a significant difference between the means of these two groups and this result suggests that this hypothesis holds.

7.1.3.3 GH3: The user has better performance in the software of Nokia smartphone if he/she has experience using Nokia mobile phone before

According to the statistical analysis, no significant difference between the means of two subject groups exists. Therefore, we could not verify this hypothesis.

7.1.3.4 GH5: The user has better performance for the task which requires user inputting characters if he/she usually sends SMS

In the scenarios of "M-Travel", the independent-samples *t* test analysis indicated that 32 subjects, who had less experience on SMS, had a mean of 417.52 and 27 subjects, who had abundant experience on SMS, had a mean of 438.85 and the means did not differ significantly at the $p < .05$ level ($p = .198$). Levene's test for Equity of Variances did show significant differences ($p < .05$). According to the above result, there was a significant difference between the means of these two groups and this result had the same implication with this hypothesis.

7.1.2.5 GH6: The user has better performance in the experiment if he/she has IT background

In the scenarios of "Movie on Phone", the independent-samples *t* test analysis indicated that 26 subjects with no IT background had a mean of 424.83 while 33 subjects with IT background had a mean of 429.21 and the means did not differ significantly at the $p < .05$ level ($p = .811$). Levene's test for Equity of Variances did show significant differences ($p < .01$). The above result shows a significant different between the means of these two groups and had a similar implication with this hypothesis.

7.2 Qualitative Analysis

During our usability experiment, other than the quantitative data, there were a lot of useful information from users, such as their comments and suggestions towards the usability of the software application. These qualitative findings not only can verify our hypotheses, but also provide valuable input reference for improvement of software design.

7.2.1 GH4: The user has better performance when using the next software application after using the previous one

Many of the subjects felt that the second software application in the experiment had better performance because they had some warm up during the first software application. Many problems had been solved after the previous application so that the next

application could be executed more smoothly. As a result, this hypothesis may hold according to their comment.

7.2.2 SH1: The colourful screen layout with graphics is more attractive than plain text layout

In our usability experiment, we found out that different users had different standard or judgment for a usable layout. Some of them preferred colourful screen layout but others liked simple layout because they had different view points and different concerns for the usage of that software application. As a result, this hypothesis might not be true in all the cases because it involved too much subjective judgment.

7.2.3 SH3: Users feel difficult to walk when they are using the application in smartphone concurrently

Although many subjects claimed that they did not feel any difficulty when they were using the software application on smartphone during their walking the observer found that most of the subjects walked slower when they were using the application compared with their normal walking pace. The observer needed to assist some subjects to walk occasionally in order to help them make a turn. As a result, the subjects should have some difficulty in this situation and this hypothesis may hold.

7.2.4 SH4: Users find it more convenient if they can use only one hand to control the application in smartphone

Most of the subjects used one hand to control the application especially during their walking. When they were using "M-Travel", some subjects used both hands to control occasionally and they explained that it was because of the design of soft-key. For example, the confirm function was the left soft-key at first and then it changed the right soft-key. They needed to put two thumbs on both left and right soft-keys in order to have a quick response. If the control flow of soft-key could be designed better the confirm function should be, for example, always on the right soft-key all the time. This could let them control the application more smoothly because the application usually requires them to press 'confirm' only. This suggests that one hand control was better than both hand control.

7.2.5 SH5: Sound effect plays a significant role in smartphone

Only few subjects agreed that sound effect was useful in this kind of software application. Some of the subjects said that they would like to turn off the sound effects because this might disturb their control. Many subjects ignored those sound effects because they had already paid so much attention on the screen. The sound effects were useless in outdoor because the background environment might be already quite noisy. Therefore, this hypothesis may not hold.

7.2.6 SH7: The usability of the software application is not satisfactory

if the number of key-pressing per task is too much

Apart from some subjects who were experts in typing characters on mobile phone because of their previous SMS typing experience, many subjects needed to put much effort to type every character on smartphone. If the string contained some capital letter or number, they needed more time to press more keys which decreased the smoothness for completing their tasks. All of them agreed that each scenario should be finished by pressing as few keys as possible. As a result, this hypothesis should hold.

7.3 Summary

Table 7.1 gives the summary of our metrics which was verified by our experiments. Column 1 provides the abbreviation of usability metrics and their names are provided in Column 2. Column 3 lists the verification method based on the experimental result.

Abb.	***Usability Metrics***	***Verification***
NSH	Number of Seek Help	Quantitative Analysis (Chi-Square)
TT	Total Time	Quantitative Analysis (Regression Test)
FE	Frustration Expression	Quantitative Analysis (Regression Test)

Table 7.1 The Usability Metric List proved by this study

Table 7.2 gives the summary of our hypothesis which was verified by our experiments. Column 1 provides the abbreviation of our hypothesis and their names are provided in Column 2. Column 3 lists the verification method based on the experimental result.

Abb.	*Hypothesis*	*Verification*
GH2	The user has better performance in the experiment if he/she has prior experience of using WAP (Wireless Application Protocol) phone or smartphone.	Quantitative Analysis (*t*-Test)
GH4	The user has better performance when using the next software application after using the previous one.	Qualitative Analysis
GH5	The user has better performance for the task which requires user inputting characters if he/she usually sends SMS.	Quantitative Analysis (*t*-Test)
GH6	The user has better performance in the experiment if he/she has IT background.	Quantitative Analysis (*t*-Test)
SH2	The processing time and response time of the application is significant for the usability of the software.	Quantitative Analysis (Regression Test)
SH3	Users feel difficult to walk when they are using the application in smartphone concurrently.	Qualitative Analysis
SH4	Users find it more convenient if they can use only one hand to control the application in smartphone.	Qualitative Analysis
SH6	If users require a lot of extra help, the usability of the software application is not satisfactory.	Quantitative Analysis (Regression Test)
SH7	The usability of the software application is not satisfactory if the number of key-pressing per task is too much.	Qualitative Analysis
SH8	Users only show their frustrated expressions which represents their dissatisfactory.	Quantitative Analysis (Regression Test)
SH9	'Step-by-Step' approach is better than 'All-in-One'.	Quantitative Analysis (*t*-Test)

Table 7.2 The Hypothesis List provided by this study

According to the result from our experiment, 75% of those tested usability metrics were verified successfully. At the same time, 73% of our hypotheses were proved by different analysis. This shows that the metrics and hypotheses we proposed in this study are some objective rules or guidelines for the software development industry because they can be proved by in the scientific way. The software developer can use our software metrics directly when they want to evaluate the usability of the software in the

smartphone platform or the environment similar to this platform. In the coming future, there should be more innovative mobile devices which our proposed usability cannot be covered. However, the developer can follow the metric conversation in this study to develop some new usability metrics in a systematical way. On the other hand, these new usability metric outputs have the same scale and meaning. This can cultivate a common way for the usability evaluation.

Chapter 8 Conclusion

The purpose of this study was to propose some usability metrics which can evaluate the software application in smartphone. At the same time, different usability issues in this platform were investigated. Also, a usability metric list has been consolidated based on a common scale. According to the results from our experiments, the design of the flow of soft key pressing played a significant role for smartphone usability and the flow of application should be designed simply. Another contribution of this research is that we have validated three of our proposed usability metrics for the application in smartphone: Number of Seeking Help (NSH), Total Time (TT) and Frustration Expression (FE). Different group of subjects, such as different age or different education background, may have different performance and then the final metrics results will also be different. However, one of the objectives of our study is to prove the validity of our proposed usability metrics. The number of subjects in this experiment was 60 which passed the magic number of statistics (30) and we did find out the correlations between the user performance and the usability metrics. Thus, these metrics are valid for different kinds of users.

The usability experiment is very important in this research. I have studied different previous research which had usability evaluation and tried to follow their steps and procedures when I was designing my own experiment. Some of these researches are from psychology researchers and these experiments were designed very well in order to avoid the disturbance to subjects. At the same time, if a subjects was disturbed, an observer could investigate during the interview session because our experiment was conducted using “Thick Aloud” mode, and had a discussion session after each small

task. Thus, the communication between a subject and an observer was very good, and the observer could find out any big mistake about disturbance.

Because of the increasing demand of software applications in smartphone, development of these applications will become a trend in software industry. Their usability is a significant factor for their success because users are not willing to use any application which is difficult to use. According to the findings of this paper, such as our proven hypotheses, developers can learn the usability issues of the software application in the smartphone environment and even other handheld devices. They can measure the usability level of their product by using our proposed usability metrics. Based on the evaluation results, the usability problems in their applications can be identified and these usability metrics can improve the usability level of software application in the smartphone platform in the future.

In our usability metric list, four types of metric classification have been identified. Software developers can check the list and identify appropriate metrics for their usability evaluation. They can easily locate metrics that can be used at a particular development stage, metrics that can be obtained under different types of evaluation, metrics that can be used for different usability attributes, and metrics that can be automated. These classifications represent the macro-view of usability testing.

This study is a stepping-stone for further research into automatic usability evaluation. For each usability attribute, some representative metrics can be chosen from the metrics list. An automatic usability evaluation tool will be built. This will save time and manpower during usability evaluation. The first step of this future work will be to

choose the metrics from Product Automation and Source Code Automation since they can be easily automated.

Another area for future study is to determine the parameters for metrics such as the Reference Value and the Standard Value. Many of these values have not been fully investigated. As this study only verified three of our proposed usability metrics, further usability test on the software application in smartphone, which should evaluate additional types of applications, can be conducted in order to verify our remaining metrics.

Publication Arising from the Thesis

Jo C. C. W. and Leung H. K. N., “A Systematic Approach of Consolidating and Computing Usability Metrics”, Proceeding of IASTED Software Engineering and Applications 2003, Marina del Rey, CA, 2003, 60-68

Reference

1. Babiker E. M., Fujihara H. and Boyle C. D. B., “A Metric for Hypertext Usability”, Proceedings of the 9th annual international conference on Systems documentation, October 1991, 95-104

2. Brewster S., Lumsden J., Bell M., Hall M. and Tasker S., “Multimodal ‘Eyes-Free’ Interaction Techniques for Wearable Devices”, Proceedings of ACM CHI 2003, Fort Lauderdale, FL, 2003, 463-480

3. Buyukkokten O., Garcia-Molina H. and Paepcke A., “Accordion Summarization for End-Game Browsing on PDAs and Cellular Phones”, Proceedings of the Conference on Human Factors in Computing System (CHI’01), 2001

4. Chan H., Lee R., Dillon T. and Chang E., “E-Commerce: Fundamentals and Application”, John Wiley & Sons Ltd, 2001

5. Chang E. J., Dillon T. S. and Cook D., “An Intelligent System Based Usability Evaluation Metric”, Intelligent Information Systems (IIS’97). IEEE Computer Society Press, 1997, 218-226

6. Constantine L. L., “Visual Coherence and Usability: A Cohesion Metric for Assessing the Quality of Dialogue and Screen Designs”, OzCHI’96 Proc. Los Alamitos, Calif.: IEEE Computer Society Press, 1996

7. Constantine L. L. and Lockwood L. A.D., “Software for Use: A Practical Guide to the Models and Methods”, ACM Press, 1999

8. Dix A., Finlay J., Abowd G. and Beale R., “Human-computer Interaction (Second Edition)”, Prentice Hall Europe, 1998

9. ISO 9126, Software Product Evaluation – Quality Characteristics and Guidelines for Their Use, 1991

10. Frøkjær E., Hertzum M. and Hornbæk K., “Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated?”, Proceedings of ACM CHI 2000, Conference on Human Factors in Computing Systems, April 2000

11. Helal S., “Pervasive Java”, Pervasive Computing, IEEE, January-February, 2002, 82-85

12. Helal S., “Pervasive Java, Part II”, Pervasive Computing, IEEE, April-June, 2002, 85-89

13. Ivory M. Y. and Hearst M. A., “The state of the Art in Automating Usability Evaluation of User Interfaces”, ACM Computing Surveys(CSUR) 33, 4, December 2001, 470-516

14. Ivory M. Y., Sinha R R. and Hearst M. A., “Empirically Validated Web Page Design Metrics”, Proceedings of ACM CHI 01, Conference on Human Factors in Computing Systems, March 2001

15. Kiili K., “Evaluating WAP Usability: “What Usability?””, Proceedings of the IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE’02), 2002

16. Kristoffersen S. and Ljungberg F., ““Making Place” to Make IT Work: Empirical Explorations of HCI for Mobile CSCW”, Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work, 1999, 331-317

17. Law L. C. and Hvannberg E. T., “Complementarity and Convergence of Heuristic Evaluation and Usability Test: A Case Study of UNIVERSAL Brokerage Platform”, Proceedings of NordiCHI 2002, October 2002

18. Masoodian M. and Lane N., “An Empirical Study of Textual and Graphical Travel Itinerary Visualization using Mobile Phones”, Proceedings of the Fourth Australian user interface conference on User interfaces 2003, Adelaide, Australia, February, 2003, 11-18

19. Myers B.A, Lie K. P. and Yang B., “Two-handed Input Using a PDA and a Mouse”, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, April 01-06, 2000, 41-48

20. Myers B. A., Nichols J., Wobbrock J. O., Litwack K., Higgins M., Hughes J., Harris T. K., Rosenfeld R., and Pignol M., "Handheld Devices for Control", Human-Computer Interaction Consortium (HCIC '03), Winter Park, CO, February 2003

21. Niklfeld G. and Pucher M., “Mobile Multi-modal Data Services for GPRS Phones and Beyond”, Proceedings of 4th IEEE Conference on Multi-model Interfaces ICMI-02, 2002

22. Noble J. and Constantine L. L., “Interactive Design Metric Visualization: Visual Metric Support for User Interface Design”, In Grundy, J., & Apperley, M.(eds.) Proceedings, Sixth Australian Conference on Computer-Human Interaction. Los Alamitos, Calif: IEEE Computer Society Press, 1996

23. Pirhonen A., Brewster S. and Holguin C., “Gestural and Audio Metaphors as a Means of Control for Mobile Devices”, Proceedings of ACM CHI 2002, Minneapolis, Minnesota, 2003, 291-298

24. Seffah A., Kececi N. and Donyaee M., “QUIM: A Framework for Quantifying Usability Metrics in Software Quality Models”, In Second Asia-Pacific Conference on Quality Software (APAQS), 2001, 311-317

25. Sgouros, N. M. and Gerogiannakis S., “Integrating WAP-based Wireless Devices in Robot Teleoperation Environments”, Proceedings of 2002 IEEE International Conference on Robotics & Automation (ICRA-02), Washington, DC, USA, 1191-1196

26. Tang J. C., Yankelovich N. and Begole J., “ConNexus to Awarenex: Extending Awareness to Mobile Users”, Proceedings of CHI 2001, Seattle, Washington, March, 2001

27. Virrantaus K., Veijalainen J., Markkula J., Katasonov A., Garmash A., Tirri H., Terziyan V., “Developing GIS-Supported Location-Based Services”, Proceedings of the First International Workshop on Web Geographical Information Systems (WGIS 2001), Kyoto, Japan, 3-6 December, 2001, 423-432

28. Weiss S., “Handheld Usability”, John Wiley & Sons Ltd, 2002

29. White J., “An Introduction to Java 2 Micro Edition (J2ME); Java in Small Things”, Proceedings of the 23rd International Conference on Software Engineering, 2001, 724-725

30. Wigdor D. and Balakrishnan R., “TiltText: Using Tilt for Text Input to Mobile Phones”, Proceedings of ACM UIST 2003, Vancouver, British Columbia, 2003

31. Wiklund M. E., “Usability in Practice: How Companies Develop User-Friendly Products”, Academic Press, 1994

32. Wobbrock J. O., Forlizzi J., Hudson S. E., and Myers B. A., “WebThumb: Interaction Techniques for Small-Screen Browsers”, Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '02). Paris, France, October 2002, 205-208

33. Wobbrock J. O., Myers B. A. and Kembel J. A., “EdgeWrite: A Stylus-Based Test Entry Method Designed for High Accuracy and Stability of Motion”, Proceedings of UIST 2003, 61-70

34. www.my-siemens.com

35. www.nokia.com

Appendix A The Specifications of Smartphone

NOKIA 7250 [35]

Size
- 105 x 44 x 19 mm, 73 cc

Weight
- 92 g (battery included)

Display
- High-resolution, passive matrix color display
- Supports 4096 colors within 128 x 128 pixels
- Up to 8 lines (Latin) / 6 lines (Chinese) in message viewing
- Adjustable display brightness control

Data Transfer
- GPRS (General Packet Radio Service)
- HSCSD (High-Speed Circuit-Switched Data)

Mobile Internet Access
- WAP 1.2.1 Browser (via GRPS or CSD)

Data exchange with PC
- IR and cable

The following features share a memory pool of 3,5 MB
- MMS messages (max size 45 kB per MMS message)
- Ringing tones in Gallery (20 preset, all removable)
- Images (10 preset, all removable)

Java™ applications (3 preset, max download size 64 kB per application)

Preinstalled Java™ Applications
- Converter II (currency, area, length, mass, temperature, and other conversions)
- Games: Triple Pop and Bounce

**All applications user removable

Shared memory

The following features in this phone may share memory: (phone book, text and multimedia messages, images and ringing tones in gallery, calendar, to-do notes, and Java games and applications). Using any such features may reduce the memory available for any features sharing memory. This is especially true with heavy use of any of the features (although some of the features may have a certain amount of memory specially allotted to them in addition to the amount of memory shared with other features).

SIEMENS S57 [34]

Size
- 101 x 42 x 18 mm, 69 cc

Weight
- 85 g (battery included)

Display
- High-resolution, passive matrix color display
- Supports 256 colors within 101 x 80 pixels
- Up to 7 lines (Latin) in message viewing

Data Transfer
- GPRS (General Packet Radio Service) Level 10

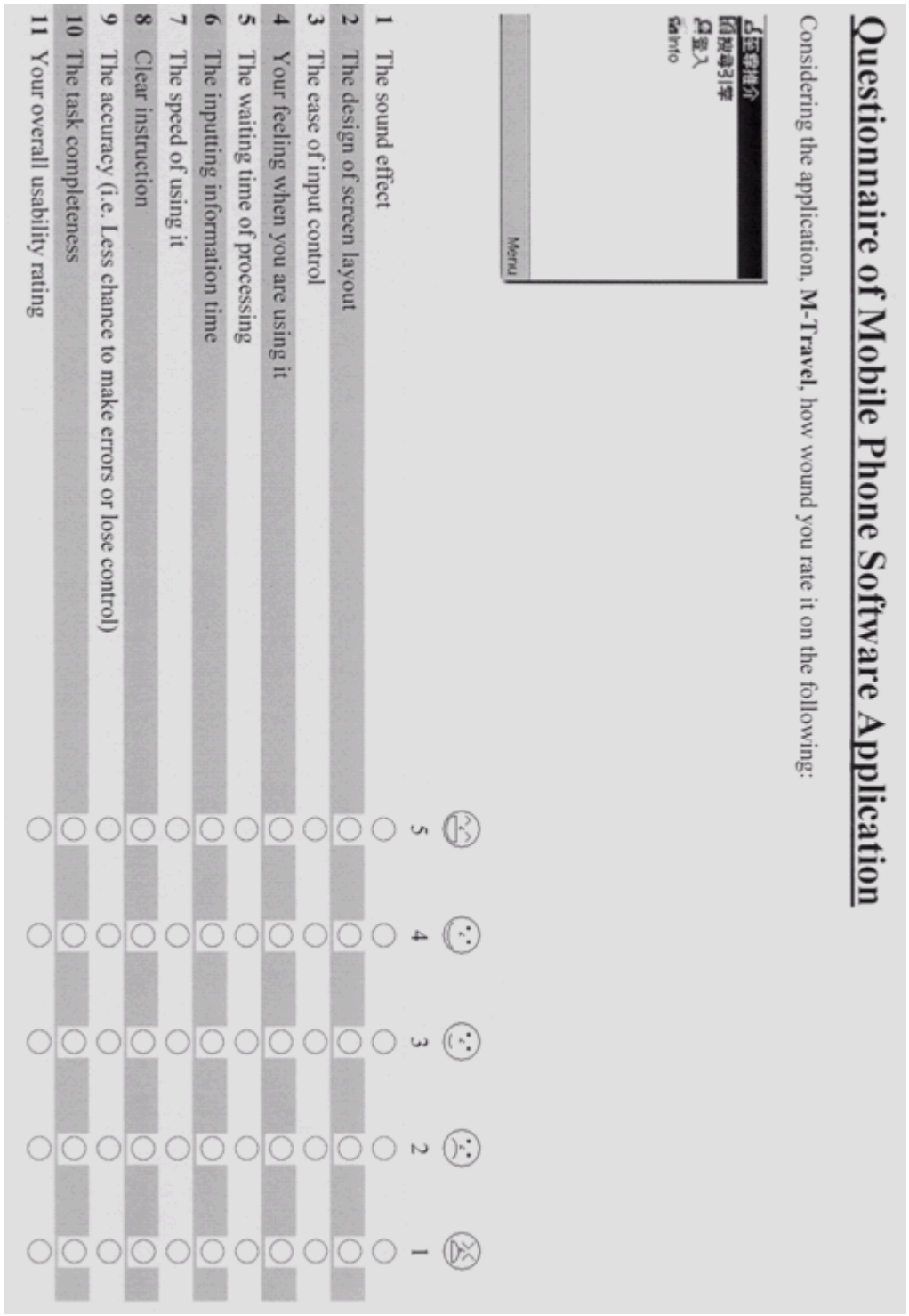
Mobile Internet Access
- WAP 1.2.1 Browser (via GRPS or CSD)

Data exchange with PC
- IR and cable

Additional Information
- Intelligent Typing (T9)
- **Softkey programmable**
- Support bluetooth
- Support JAVA™ applications
- Support JAVA™ games

Appendix B The Questionnaire Samples

The questionnaire to evaluate the software M-Travel



Questionnaire of Mobile Phone Software Application

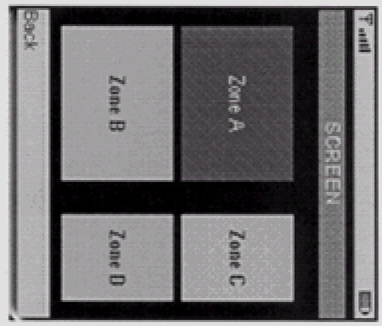
Considering the application, **M-Travel**, how wound you rate it on the following:

		5	4	3	2	1
1	The sound effect	○	○	○	○	○
2	The design of screen layout	○	○	○	○	○
3	The ease of input control	○	○	○	○	○
4	Your feeling when you are using it	○	○	○	○	○
5	The waiting time of processing	○	○	○	○	○
6	The inputting information time	○	○	○	○	○
7	The speed of using it	○	○	○	○	○
8	Clear instruction	○	○	○	○	○
9	The accuracy (i.e. Less chance to make errors or lose control)	○	○	○	○	○
10	The task completeness	○	○	○	○	○
11	Your overall usability rating	○	○	○	○	○

The questionnaire to evaluate the software Movie on Phone

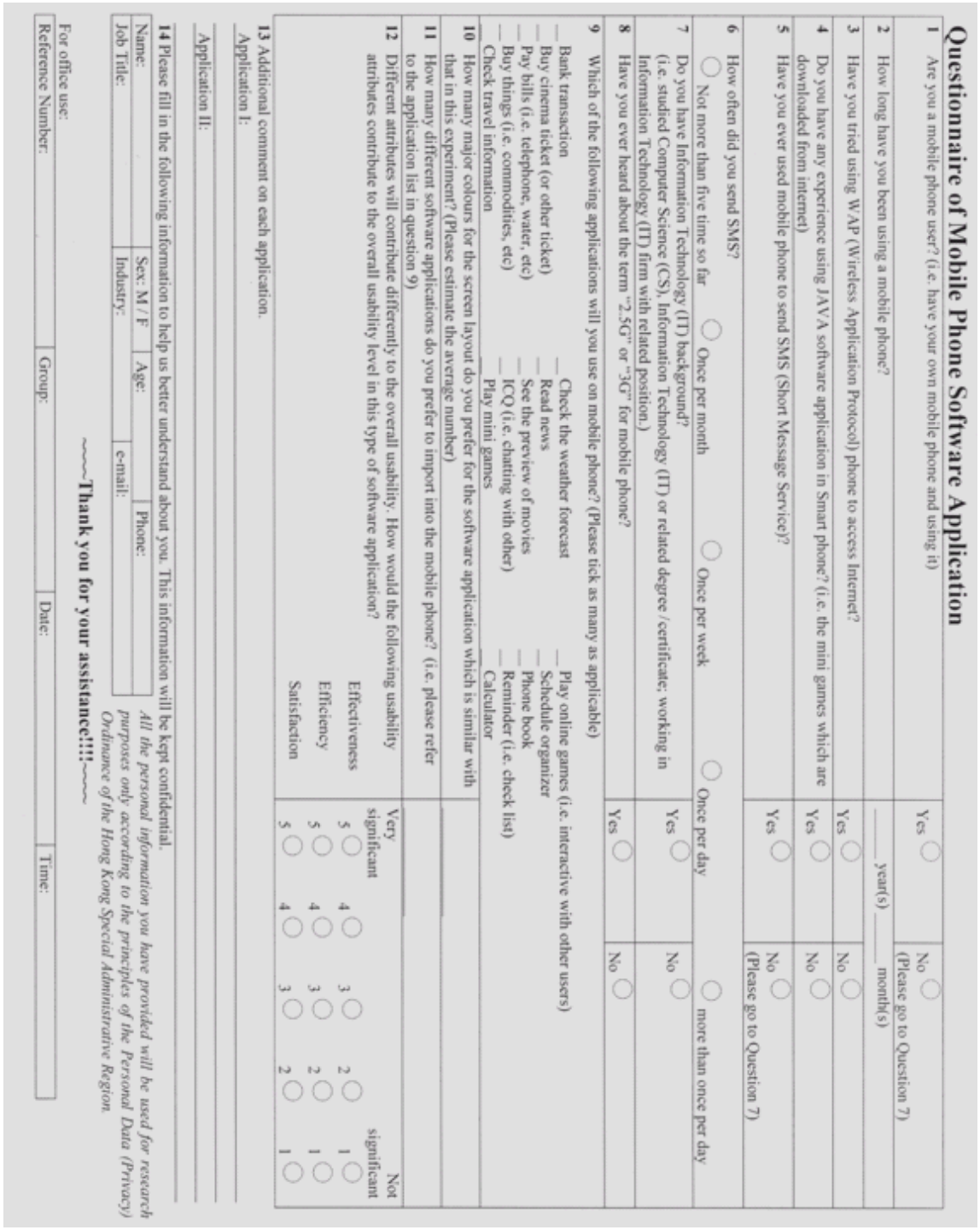
Questionnaire of Mobile Phone Software Application

Considering the application, **Movie on Phone**, how wound you rate it on the following:



		5	4	3	2	1
1	The sound effect	○	○	○	○	○
2	The design of screen layout	○	○	○	○	○
3	The ease of input control	○	○	○	○	○
4	Your feeling when you are using it	○	○	○	○	○
5	The waiting time of processing	○	○	○	○	○
6	The inputting information time	○	○	○	○	○
7	The speed of using it	○	○	○	○	○
8	Clear instruction	○	○	○	○	○
9	The accuracy (i.e. Less chance to make errors or lose control)	○	○	○	○	○
10	The task completeness	○	○	○	○	○
11	Your overall usability rating	○	○	○	○	○

The questionnaire to evaluate users background



Questionnaire of Mobile Phone Software Application

1 Are you a mobile phone user? (i.e. have your own mobile phone and using it) Yes ○ No ○ (Please go to Question 7)

2 How long have you been using a mobile phone? _____ year(s) _____ month(s)

3 Have you tried using WAP (Wireless Application Protocol) phone to access Internet? Yes ○ No ○

4 Do you have any experience using JAVA software application in Smart phone? (i.e. the mini games which are downloaded from internet) Yes ○ No ○

5 Have you ever used mobile phone to send SMS (Short Message Service)? Yes ○ No ○ (Please go to Question 7)

6 How often did you send SMS?
○ Not more than five time so far ○ Once per month ○ Once per week ○ Once per day ○ more than once per day

7 Do you have Information Technology (IT) background? (i.e. studied Computer Science (CS), Information Technology (IT) or related degree /certificate; working in Information Technology (IT) firm with related position.) Yes ○ No ○

8 Have you ever heard about the term "2.5G" or "3G" for mobile phone? Yes ○ No ○

9 Which of the following applications will you use on mobile phone? (Please tick as many as applicable)

__ Bank transaction	__ Check the weather forecast	__ Play online games (i.e. interactive with other users)
__ Buy cinema ticket (or other ticket)	__ Read news	__ Schedule organizer
__ Pay bills (i.e. telephone, water, etc)	__ See the preview of movies	__ Phone book
__ Buy things (i.e. commodities, etc)	__ ICQ (i.e. chatting with other)	__ Reminder (i.e. check list)
__ Check travel information	__ Play mini games	__ Calculator

10 How many major colours for the screen layout do you prefer for the software application which is similar with that in this experiment? (Please estimate the average number)

11 How many different software applications do you prefer to import into the mobile phone? (i.e. please refer to the application list in question 9)

12 Different attributes will contribute differently to the overall usability. How would the following usability attributes contribute to the overall usability level in this type of software application?

	Very significant						Not significant
Effectiveness	5 ○	4 ○	3 ○	2 ○	1 ○		
Efficiency	5 ○	4 ○	3 ○	2 ○	1 ○		
Satisfaction	5 ○	4 ○	3 ○	2 ○	1 ○		

13 Additional comment on each application.
Application I:

Application II:

14 Please fill in the following information to help us better understand about you. This information will be kept confidential.

Name:	Sex: M / F	Age:	Phone:
Job Title:	Industry:		e-mail:

All the personal information you have provided will be used for research purposes only according to the principles of the Personal Data (Privacy) Ordinance of the Hong Kong Special Administrative Region.

~~~~**Thank you for your assistance!!!**~~~~

For office use:

Reference Number:	Group:	Date:	Time:
~~~~

Appendix C The Scenario Sheet

The Scenario Sheet for M-Travel

Application: M-Travel
SIMENS S-57

THE HONG KONG
POLYTECHNIC UNIVERSITY
Department of Computing
By Craig Jo

Scenario for smartphone experiment

Scenario One	Familiar with the application	
Task One: 1, Enter this application 2, Browse the application freely to familiar the function in this application 3, If you access some parts which need you to login, please leave that part and browse the other part. 4, You have 2 minutes for this task		

Scenario Two	Login to the system	
Task One: 1, Go to the main menu of this application 2, Select the function '登入' 3, Your login is 'john' 4, Input your login in the field '用戶名稱.' *(Let me know if you have already finished this task)*		
Task Two: 1, Your Password is 'apzerklq' 2, Input your Password in the field '密碼' 3, Then login to this application *(Let me know if you have already finished this task)*		

Scenario Three	Reserve tour	
Task One: 1, Select the function '搜尋引擎' 2, You need to search the tour has the price within $500 to $2000 3, Your tour starting date is after 01/09/2003 3, Use appropriate search method and input this criteria 4, Found out the search function inside 選項 and select it. *(Let me know if you have already finished this task)*		
Task Two: 1, The appropriate result will be listed out. 2, Find out which tour is the most expensive for going to 杭州 3, Read the detail information of this tour *(Let me know if you have already finished this task)*		
Task Three: 1, In the same page, select 天氣報告 to check the four season weather of 杭州 2, Reserve 5 seats for this tour 3, Then try to reserve 4 seats from the page of the tour detail information 4, Press the '留位' function and read the detail information in the confirmation page and then confirm your reservation. *(Let me know if you have already finished this task)*		

The Scenario Sheet for Mobile on Phone

Application: Movie on Phone
Nokia 7250

THE HONG KONG
POLYTECHNIC UNIVERSITY
Department of Computing
By Craig Jo

Scenario One	Familiar with the application	
Task One: 1, Enter this application 2, Browse the application freely to familiar the function in this application 3, If you access some parts which need you to login, please leave that part and browse the other part. 4, You have 2 minutes for this task		

Scenario Two	Login to the system	
Task One: 1, Go to the main menu of this application 2, Select the function 'Buy Tickets' 3, Your login is 'Susan' 4, Input your login in the field 'User ID.' *(Let me know if you have already finished this task)*		
Task Two: 1, Your Password is 'wz3fbc' 2, Input your Password in the field 'Password' 3, Then login to this application *(Let me know if you have already finished this task)*		

Scenario Three	Reserve Ticket	
Task One: 1, You are now in the function ‘Buy Tickets’ 2, You are deciding to buy which time slot of ticket by search the cinema 3, Kowloon Bay is your favour cinema 3, You plan to see the movie which calls ‘The Recruit’ 4, Your available time is on 31-8-2003 **after** 7pm 5, Base on the above criteria, select the appropriate choice in the application *(Let me know if you have already finished this task)*		
Task Two: 1, The seating plan will be shown 2, You want to find a seat in Zone A 3, You will buy 4 tickets and they are located from C3 to C6 4, Click the select button at C3 first, then use arrow key to reserve more than one seat *(Let me know if you have already finished this task)*		
Task Three: 1, Then click the ‘Buy’ button to reserve the seats 2, Now the confirmation information of your reservation will be shown 3, Check whether the information of your tickets is correct or not 4, Confirm your reservation if all the things are OK. *(Let me know if you have already finished this task)*		

Appendix D Statistical Analysis Result

t-Test (Total Mark of M-Travel [ALLT] VS Total Mark of Movie on Phone [ALLM])

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
ALLT	396.2084	59	71.94446	9.36637
ALLM	427.2795	59	72.43729	9.43053

Paired Samples Correlations

	N	Correlation	Sig.
ALLT & ALLM	59	.398	.002

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
ALLT - ALLM	-31.0711	79.18689	10.30926	-51.7074	-10.4349	-3.014	58	.004

Chi-Square Test (Result of “Number of Seeking Help” in Secenario 2 VS User Preference)

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	12.082	4	.017
Likelihood Ratio	12.804	4	.012
Linear-by-Linear Association	6.353	1	.012
N of Valid Cases	56		

a 4 cells (44.4%) have expected count less than 5. The minimum expected count is 1.50.

Chi-Square Test (Result of “Number of Seeking Help” in Secenario 3 VS User Preference)

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	11.970	4	.018
Likelihood Ratio	12.462	4	.014
Linear-by-Linear Association	.341	1	.559
N of Valid Cases	58		

a 4 cells (44.4%) have expected count less than 5. The minimum expected count is 1.33.

Chi-Square Test (Result of “Total Time” in Secenario 2 VS User Preference)

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8.805	4	.066
Likelihood Ratio	8.916	4	.063
Linear-by-Linear Association	5.910	1	.015
N of Valid Cases	57		

a 5 cells (55.6%) have expected count less than 5. The minimum expected count is 1.23.

Chi-Square Test (Result of “Total Time” in Secenario 3 VS User Preference)

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4.058	4	.398
Likelihood Ratio	3.919	4	.417
Linear-by-Linear Association	.395	1	.530
N of Valid Cases	59		

a 4 cells (44.4%) have expected count less than 5. The minimum expected count is 1.49.

Regression Test (Result of “Total Time” for Movie on Phone [TTLTIMEM] VS Questionnaire Result for Movie on Phone [inputting time M, speed of use M, overall rating M])

Correlations

		inputting time M	speed of use M	overall rating M
Pearson Correlation	TTLTIMEM	.338	.344	.277
Sig. (1-tailed)	TTLTIMEM	.004	.004	.016
N	TTLTIMEM	60	60	60

Regression Test (Result of “Frustration Expression” for M-Travel [TTLTIMET] VS Questionnaire Result for M-Travel [screen layout T, input control T, feeling T, overall rating T])

Correlations

		screen layout T	input control T	feeling T	overall rating T
Pearson Correlation	TTLFRUST	.285	.276	.359	.204
Sig. (1-tailed)	TTLFRUST	.014	.016	.002	.059
N	TTLFRUST	60	60	60	60

Regression Test (Result of “Frustration Expression” for Movie on Phone [TTLTIMEM] VS Questionnaire Result for Movie on Phone [screen layout M, input control M, feeling M, overall rating M])

Correlations

		screen layout M	input control M	feeling M	overall rating M
Pearson Correlation	TTLFRUSM	.275	.258	.323	.299
Sig. (1-tailed)	TTLFRUSM	.017	.023	.006	.010
N	TTLFRUSM	60	60	60	60

t-Test (Questionnaire [WAP phone experiment] VS Total Metric Mark [OVERALLT for M-Travel, OVERALLM for Movie on Phone])

Group Statistics

	WAP phone experiment	N	Mean	Std. Deviation	Std. Error Mean
OVERALLT	No	17	376.6353	48.06452	11.65736
	Yes	43	403.4128	77.92360	11.88324
OVERALLM	No	16	407.7550	47.19429	11.79857
	Yes	43	434.5440	79.06606	12.05746

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
OVERALLT	Equal variances assumed	10.751	.002	-1.317	58	.193	-26.7775	20.32767	-67.46774	13.91275
	Equal variances not assumed			-1.609	47.138	.114	-26.7775	16.64648	-60.26330	6.70831
OVERALLM	Equal variances assumed	4.593	.036	-1.270	57	.209	-26.7890	21.10174	-69.04446	15.46655
	Equal variances not assumed			-1.588	45.117	.119	-26.7890	16.86976	-60.76396	7.18605

t-Test (Questionnaire [SMS] VS Total Metric Mark [OVERALLT for M-Travel, OVERALLM for Movie on Phone])

Group Statistics

	SMS	N	Mean	Std. Deviation	Std. Error Mean
OVERALLT	No	32	384.6663	60.30231	10.66004
	Yes	28	408.5796	81.52701	15.40716
OVERALLM	No	32	417.5172	79.77584	14.10251
	Yes	27	438.8489	62.13669	11.95821

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
OVERALLT	Equal variances assumed	4.986	.029	-1.302	58	.198	-23.9134	18.36700	-60.67893	12.85214
	Equal variances not assumed			-1.276	49.215	.208	-23.9134	18.73545	-61.55953	13.73274
OVERALLM	Equal variances assumed	2.516	.118	-1.130	57	.263	-21.3317	18.88438	-59.14703	16.48362
	Equal variances not assumed			-1.154	56.672	.253	-21.3317	18.48999	-58.36190	15.69850

t-Test (Questionnaire [IT background] VS Total Metric Mark [OVERALLT for M-Travel, OVERALLM for Movie on Phone])

Group Statistics

	IT background	N	Mean	Std. Deviation	Std. Error Mean
OVERALLT	No	26	400.6919	65.87656	12.91946
	Yes	34	392.1047	76.10743	13.05232
OVERALLM	No	26	424.8308	54.68875	10.72535
	Yes	33	429.2082	84.64788	14.73530

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
OVERALLT	Equal variances assumed	1.011	.319	.459	58	.648	8.5872	18.72559	-28.89613	46.07056
	Equal variances not assumed			.468	57.052	.642	8.5872	18.36506	-28.18745	45.36189
OVERALLM	Equal variances assumed	8.008	.006	-.229	57	.820	-4.3774	19.15246	-42.72955	33.97472
	Equal variances not assumed			-.240	55.095	.811	-4.3774	18.22532	-40.90036	32.14553

Appendix E Statistical Raw Data

Data from Experimental Tasks (M-Travel)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	One	No	No	Yes		No	No	NA	No	No	No	No	Yes	Words should be smaller, show more information	No
3	Two	No	Yes	No		No	No	Yes	Yes	No	No	No	No	Many information on the screen	No
4	One	No	No	Yes		No	No	Yes	Yes	No	No	Yes	No	show enough information	No
5	One	No	Yes	Yes		No	No	Yes	Yes	Yes	No	No	No		No
6	One	Yes	No	Yes		No	No	Yes	Yes	No	No	No	No	need wall paper, not enough picture	No
7	One	Yes	No	No		No	No	Yes	No	Yes	No	No	No	not easy to be bored, good feeling	No
8	One	No	No	No		Yes	No	Yes	Yes	No	No	Yes	No	one screen can show all option	No
9	One	No	No	No		Yes	No	Yes	Yes	Yes	No	No	No		No
10	One	No	Yes	No		No	Yes	Yes	Yes	No	Yes	No	No	Simple	No
11	Two	No	Yes	No		No	Yes	Yes	Yes	Yes	Yes	No	No		No
12	Two	No	No	No	Not famous	No	No	Yes	No	Yes	No	Yes	No		No
13	Two	No	No	Yes		No	Yes	Yes	Yes	Yes	Yes	No	No	Chinese	No
14	Two	No	Yes	No		No	No	Yes	Yes	No	No	No	No	Word is too big, usually need to scroll	No
15	One	No	No	No	need time for thinking, no need too fast	No	Yes	No	Yes	No	No	No	No		No
16	Two	No	No	No	Dont have any target	No	No	NA	No	No	No	Yes	No		No
17	One	No	Yes	No		No	Yes	Yes	Yes	No	No	Yes	No	word is too large	No
18	One	No	No	Yes		No	Yes	Yes	Yes	No	No	Yes	No		No
19	Two	No	No	No	quicker	No	No	Yes	No	Yes	No	No	No		No
20	One	No	Yes	No		No	Yes	Yes	Yes	Yes	No	No	No	movie	No
21	One	No	No	Yes		No	No	Yes	No	No	No	No	No	softkey layout have problem	No
22	One	No	No	Yes		No	Yes	Yes	Yes	No	Yes	No	No		No
23	Two	No	No	No	Not famous since it is testing	No	Yes	NA	Yes	No	No	No	Yes	Too much information	No
24	One	No	No	No		Yes	Yes	Yes	Yes	No	Yes	No	No	Word is bigger	No
25	One	No	Yes	Yes		No	Yes	Yes	Yes	No	No	Yes	No		No
26	Two	No	No	No	softkey is no good	No	Yes	NA	No	No	No	No	No	too complex	No
27	Two	No	No	Yes		No	Yes	NA	Yes	Yes	No	No	No		No
28	One	Yes	No	No		No	No	Yes	Yes	No	Yes	No	No		No
29	One	No	No	No		Yes	Yes	Yes	No	Yes	No	No	No		No
30	One	No	No	No		No	Yes	No	No	No	No	Yes	No		No
31	Two	No	Yes	No		No	Yes	NA	No	No	No	No	Yes		No
32	One	No	No	Yes		No	No	Yes	Yes	No	Yes	No	No		No
33	One	No	No	No		Yes	Yes	Yes	No	No	Yes	No	No	Chinese	No
34	Two	No	No	No	not famous	No	Yes	Yes	No	No	No	Yes	No		No
35	Two	No	No	No	Other hand no use	No	NA	Yes	Yes	No	No	Yes	No		No
36	One	No	Yes	No		No	No	Yes	No	No	No	Yes	No		No
37	Two	No	No	Yes		No	Yes	Yes	No	No	No	Yes	No		No
38	One	Yes	No	No		No	No	Yes	No	No	No	Yes	No		No
39	One	No	No	No		No	Yes	Yes	Yes	No	No	No	No	user-friendly	No
40	One	No	Yes	No		No	No	Yes	No	No	No	No	Yes		No
41	One	No	No	No		Yes	Yes	Yes	No	No	No	No	No	not clear	No
42	One	No	No	Yes		No	Yes	Yes	Yes	No	No	Yes	No		No
43	One	No	Yes	No		No	Yes	Yes	No	No	No	Yes	No		No
44	One	No	No	No	enough	No	Yes	Yes	No	No	Yes	No	No		No
45	One	No	Yes	No		No	No	Yes	Yes	Yes	No	No	No		No
46	One	No	Yes	No		No	No	Yes	No	No	No	Yes	No		No
47	One	No	Yes	No		No	Yes	Yes	Yes	Yes	Yes	No	No		No
48	One	No	Yes	No		No	Yes	Yes	No	No	Yes	No	No	Colour not sharp	No
49	One	No	No	No	One hand is enough	No	No	Yes	No	Yes	No	No	No		No
50	One	Yes	No	No		No	No	Yes	Yes	No	No	No	No	colourful, word bigger	No
51	One	No	Yes	No		No	Yes	Yes	Yes	Yes	Yes	No	No		No
52	One	No	No	No	two hands is difficult to use	No	Yes	Yes	Yes	No	Yes	No	No		No
53	Two	No	Yes	No		No	Yes	No	No	No	Yes	No	No		No
54	Two	No	Yes	No		No	Yes	Yes	No	No	No	Yes	No		No
55	Two	No	Yes	No	not famous	No	Yes	Yes	Yes	Yes	No	No	No		No
56	One	No	Yes	No		No	No	Yes	No	No	No	Yes	No		No
57	One	Yes	Yes	No		No	No	Yes	Yes	No	Yes	No	No	word is bigger	No
58	Two	No	No	No	not famous	No	Yes	Yes	Yes	No	Yes	No	No		No
59	One	No	Yes	No		No	Yes	No	Yes	Yes	Yes	No	No		No
60	One	No	No	Yes		No	Yes	Yes	No	No	Yes	No	No		No
61	Two	Yes	No	No		No	No	Yes	No	Yes	No	No	No		No

17	18	19	20	21
Softkey confuse -> dont't know how to exit	1	1	3	1
	2	0	1	0
concern about waiting time, difficult to exit because of softkey	2	1	0	1
more clear because chinese, menu is not detail enough	2	1	1	0
The flow of soft key is no good, should have enter and exit only	1	1	3	0
Good interface, Chinese is good	2	1	0	1
	2	0	0	0
attractive colour screen, slow speed, the lable of softkey is not consistant	2	0	1	0
Usually press wrong softkey because confirm key is at right	3	1	2	0
	2	0	2	0
Very slow, not enough function, most of his focus pn eyes, no need ears	1	1	1	0
	0	0	1	1
Word not clear, colour too dim	3	1	2	0
Presentation of Date is not clear (UK VS US)	2	2	1	1
	0	0	0	0
Softkey problem (the sequence of left and right)	3	2	1	0
Try to use short cut (Press '3' to select third item)	1	1	1	0
	0	0	0	0
	2	0	1	0
	1	3	0	1
	1	0	0	0
Too much information on the screen	1	0	0	0
	1	1	2	1
Softkey have problem	4	0	2	1
	1	0	0	1
Need some time to familiar	2	0	2	0
Softkey is complex	6	0	3	0
	2	0	0	0
Speed is slow, too complicate for one hand usage	1	0	0	0
menu bar is no good	5	1	1	0
	2	0	0	0
	0	1	0	0
Very slow, difficult to use	4	0	1	1
Softkey have problem -> about left and right sequence	2	1	1	0
Difficult to 'back'	1	0	2	1
Softkey usually change the flow (left, right), colourful	3	2	0	0
Easy to press wrong key, troublesome	1	1	1	1
Left-right reverse confirm key, no good for nokia user l faster	2	0	0	0
	2	0	1	1
difficult to back, 'softkey' have problem	2	1	1	0
	0	0	0	0
	0	0	0	0
	1	2	0	1
Softkey difficult to use	2	0	1	1
Slow l Word is ugly l softkey is messy	1	0	0	0
	1	2	0	0
	0	0	1	0
Softkey keep on changing left and right	2	1	2	0
	0	0	1	0
	2	1	1	0
Feel very slow	3	1	0	0
	2	2	1	0
	1	1	1	0
	0	0	0	0
	2	4	0	1
Like the process page	1	0	1	0
Feel quite slow	1	1	1	0
	2	0	1	1
Slow, softkwy have problem	1	2	0	0
	1	1	2	1

22	23	24	25	26	27
Don't show what is the input method at the begining I expect to press 'enter' key rather than scroll down after inputted login	0	2	3	0	2
Character selection bar is useful	0	0	1	1	0
Feel difficult for T9 since lack to instruction to show the input method I Password should be in the same page as login I expect to press 'enter' after login	1	2	4	0	1
Character selection list can give a good instruction	2	0	1	2	2
softkey lable make confuse	1	4	1	1	2
Input method should show on screen	0	2	0	0	2
Should not use T9 for default input method, should show what input is using	0	2	1	0	2
There is some problem for T9, the character selection list is not necessary	3	0	1	1	2
T9 input have some problem, should not be placed at the begining, feel troublesome I Character selection menu is good	0	5	0	1	2
Dont show what is the input method at the begining	1	2	0	1	1
Should know what input method at the begining	0	1	1	1	2
Quite troublesome since too much input method I Mislead, default input method should be simple English	0	2	1	0	2
Easy to input wrongly for all phone I Password change to * is quite confuse	0	2	3	0	2
Very good for input selection	2	2	4	0	2
Word is very small, quite complex to input one character	1	0	2	1	1
Default input should be "ABC", input method is too messy, password should be number only	1	3	2	0	1
Should indicate what input method are using at the begining	0	1	2	1	2
Input method should not default to T9	0	1	2	0	0
Dont clear instruction for changing input method	1	1	3	0	1
Dont know what input method when using	0	0	0	0	1
Dont know which kind of input method at the begining	0	1	3	0	0
Cant expect where to input password after inputing login	0	3	2	0	0
Character selection bar is good, it is good since have sound effect	0	0	2	0	2
Expect to have an 'enter' key after login, cant expect need to scroll down	2	0	2	1	1
Dont know password is below, Character selection bar is good	0	1	1	0	1
Dont know how to change input method	0	0	1	0	0
Character selection menu is good I Softkey design flow is no good, make delay	0	1	0	0	1
Character selection menu is good	1	1	2	0	0
'T9' have some problem	1	2	5	0	1
Character selection menu is useful	0	1	1	0	1
Need to find each character, input is quite messy, character selection bar is good	0	0	0	0	1
Can't expect the password is located at the same page with login	0	2	2	0	1
Difficult to use, not famous	0	1	2	0	1
slow	3	1	2	0	0
Character selection bar is good design	1	0	2	1	1
Softkey problem	1	1	1	0	1
Very troublesome, cant expect the password input in below page	0	1	1	0	1
Dont know what input method at the begining I Character selection bar make more confuse	0	1	2	0	1
	2	4	1	1	1
Character selection bar is good	2	2	3	0	1
Dont know which input method is using I When input password, input method change suddently I assume password should be in other page	0	3	0	0	1
Dont know the input method at the beginning	0	1	0	0	2
Dont know how to use T9 at the beginning I Cant expect password should scroll down	0	0	1	0	0
Character selection bar is good	1	0	0	1	1
Cant predict where is password	0	1	4	0	1
More clear, sound effect can help	1	3	0	0	1
	2	1	0	0	1
Default input should be English I expect to press 'enter' after login rather than scroll down	0	0	0	0	1
Character selection bar is good, easy to select the char	0	2	6	0	1
Dont know which input method is using	0	0	1	0	1
Should have 'enter' after input login, dont know the place for password	0	3	1	0	0
Very troublesome to input something I Dont know password is below	0	1	1	0	1
Dont know using T9 at the begining	1	0	1	2	1
More simple	0	0	1	0	2
Very difficult to use, softkey mislead I dont know the default input method	0	0	1	0	0
Dont know T9 I Character Selection bar is good	0	1	1	0	1
Dont know what is the inpit method at the beginning	0	0	0	0	2
Character selection bar is good	0	0	3	0	1
Very slow, need to wait char by char	1	4	1	1	2
Too many input method, quite troublesome	0	0	1	0	0

28	29	30	31	32	33
3	102	45	37	192	125
0	56	25	54	20	56
2	17	29	54	128	178
1	56	58	96	95	150
1	121	50	86	118	125
0	65	38	52	90	71
0	48	31	52	114	90
2	87	54	66	110	161
0	90	50	30	126	54
1	67	51	55	80	110
4	40	55	60	120	143
0	104	38	20	263	146
0	109	47	90	195	216
1	94	51	95	121	175
1	39	48	54	129	174
2	82	41	56	95	148
2	62	79	94	170	267
1	33	27	43	57	154
1	55	49	54	90	146
2	53	132	70	118	150
0	61	33	33	67	95
1	45	32	33	78	120
3	97	100	96	151	277
1	150	46	63	118	117
0	35	47	47	94	65
1	96	36	34	87	138
1	118	48	52	140	155
2	44	48	57	96	128
1	48	40	43	90	150
1	128	43	54	131	113
1	37	27	52	125	196
1	25	57	49	124	129
1	76	56	58	90	133
1	67	58	80	125	119
3	65	92	99	161	320
0	41	37	45	54	89
1	77	79	58	82	132
1	50	35	60	96	121
1	71	46	92	115	117
1	54	74	51	121	159
0	21	44	35	119	71
1	68	30	38	93	94
1	40	58	58	78	135
1	117	62	113	100	225
1	28	19	27	57	88
1	89	87	102	181	127
1	86	57	135	114	156
1	129	78	52	99	144
6	33	32	37	75	.
0	60	53	47	136	75
1	50	65	37	118	96
1	49	78	55	152	135
1	51	44	60	139	120
1	45	33	63	110	91
0	35	180	91	66	86
1	76	42	50	108	136
2	83	133	47	104	158
2	66	40	48	101	176
1	38	57	74	149	173
1	119	101	53	71	147

34
The time delay make the software too slow to response
Clear instruction, more detail
Need some time to warm up, feel more clear, may be since the interface
After enter the item detail, when we go out, then back to item list, should be at the same item
After enter the detail of an item, when we back, should be at the item in the list I softkey design is no good
Should scroll down rather than softkey for next page I it is straight forward
User friendly except the label of softkey
Complex, softkey is quite confuse I should be scroll down for next page
'Reserve sit' is quite misleaded because dont expect to scoll doen to input the sit number
Next page should use 'scroll down' rather than softkey I soft-key have same problem
Clear, but the softkey label is different with nokia I sould effect have some help, have instruction during waiting
Softkey mislead, instruction not clear I Suggestion: change colour for instruction
Very tired to use it, response time for each key is slower I Control of softkey have problem (cant epect which function in each softkey) I only use up and down key for next page, cant expect using soft key
Dont understand how to use, easy to get loss, screen is too small, quite slow
important thing should be place at the begin, softkey is not consistant
The style of the date layout I should important data at the begining I softkey is quite messy, instruction is not clear
Screen is too small so that it show too less information I Response time is slow I the input location for reserve sit number should be changed
softkey is messy, no need use 'next pave' function, just scroll is OK
For the reserve sit, that page should not input anything because it is informative thing
Softkey is not good, flow of control is not good also. I cant predict using softkey to go next page I input and display page should not be mixed
Dont know how to go back I 'Next page' have problem I Cant expect input the sit number in the detail page
Cant reserve sit easily because dont know the location I It can show all the information before make the decision
Dont have 'scroll' concept, response time is slow
'Next page' have problem I this time is better because it is the 2nd time to do similar task
Feel troublesome for softkey (cant directly confirm)
Softkey flow is no good, in softkey selection menu, item position should be good design I should not have second page, use one page to list
Cant predict how to input 'Reserve sit' number, Cant find page number at the begining, Softkey design is no good, left right usually change
Cant predict how to use 'reserve sit'
Data showing dont mix with data input, data input should not put at last I softkey is messy
Response time is slow I one screen can only show limited detail
Cant expect how to see weather repost I softkey is messy
Response time is slow
Softkey Problem
Very slow, the input for sit cant expect, dont smooth to control
Reserve sit have problem
Cant expect how to use 'next page' function I 'Reserve sit': the function is for confirm sit rather than reserve
Should have a direct softkey for selection/confirm I cant predict where to input no. of reserve sit
Softkey problem -> easy to get loss I reserve sit -> dont know it should be scroll down
Design of softkey have problem
Cant predict how to do next page and reserve sits
Reserve sit -> cant expect to fill the number at that place I slow
Softkey design have problem
Cant expect how to use 'next page' I cant expect need to input reserve sit number at that page I softkey label is mislead
Softkey label is confusing, Softkey is messy
Dont know how to use 'Next Page' function, should just scroll down I Reserve sit number should not place at the informative page
Dont expect need to input something in a informative detail page I Next Page function should not be used, just scroll down is enough
Reserve sit have problem, softkey have problem
Cant expect need to use 'next page'
Troublesome, quite slow for response time I should be step by step (read, then input)
Reserve sit -> dont know at that location I softkey is confusing
Softkey reverse -> not user friendly I Next page function -> no good
More difficult to use in next page and weather report
After going inot detail, then go out, go to the first item rather than selected item
Reserve sit have problem -> cant expect input there
Less data, less choice, not safe I Reserve sit -> dont know should be input at that place
Page 1, 2 system is no good, should be just scroll down I Cant expect reserve sit number in that informative page
Softkey is usually change, if do something wrong, need to redo again I Next page -> should just scroll down
No need to wait so long

Key of each column

1	Ref. No.
2	How many Hands T
3	Convenient T
4	Habit T
5	Phone Shape T
6	Other T
7	NA T
8	difficult walking T
9	can one hand control T
10	sound effect (hints) T
11	Pretty T
12	Clear T
13	Acceptable T
14	Not Acceptable T
15	Other T
16	NA T
17	Remark 1 T
18	Frustration S2T1 T
19	Frustration S2T2 T
20	Extra Help S2T1 T
21	Extra Help S2T2 T
22	String Input T
23	Frustration S3T1 T
24	Frustration S3T2 T
25	Frustration S3T3 T
26	Extra Help S3T1 T
27	Extra Help S3T2 T
28	Extra Help S3T3 T
29	Time1(S2)T
30	Time2(S2)T
31	Time1(S3)T
32	Time2(S2)T
33	Time3(S2)T
34	Overall Control T

Data from Experimental Tasks (Movie on Phone)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	One	No	No	No		Yes	Yes	Yes	No	No	No	Yes	No		No
3	Two	No	Yes	No	can press quicker	No	No	Yes	No	No	Yes	No	No	easy to use	No
4	One	No	No	No		Yes	No	Yes	No	No	No	No	No	better for usage	No
5	One	No	Yes	Yes		No	Yes	NA	Yes	No	No	No	No	Complicate, not easy to use	No
6	One	No	No	No		Yes	No	NA	Yes	No	No	No	No	word make comfortable	No
7	One	Yes	No	No		No	No	Yes	No	No	No	Yes	No	should become more colourful	No
8	One	No	No	No		Yes	Yes	Yes	No	No	No	No	No	The word is too much and not tidy	No
9	One	No	No	No		Yes	Yes	Yes	No	No	Yes	No	No	lack of picture	No
10	One	No	Yes	No		No	No	Yes	No	No	Yes	No	No	Seem only just have black and white	No
11	Two	No	Yes	No		No	Yes	Yes	Yes	No	Yes	No	No	Should have wall paper, more graphic	No
12	One	No	No	No	one thumb can cover all the bottom	No	No	Yes	No	No	Yes	No	No		No
13	One	No	Yes	No		No	No	Yes	NA	No	Yes	Yes	No	too simple	No
14	Two	No	Yes	No		No	No	Yes	No	No	Yes	No	No	Consistent because all are English	No
15	One	No	No	No		Yes	No	Yes	No	No	No	No	Yes	lack of graphic	No
16	Two	No	No	No	Dont have target	No	No	Yes	No	No	No	No	No	word too small, not colourful	No
17	One	No	Yes	No		No	Yes	Yes	No	No	Yes	No	No		No
18	One	No	No	No	the situation	No	No	Yes	No	No	No	Yes	No		No
19	Two	No	No	Yes		No	No	NA	No	No	No	No	Yes		No
20	One	No	Yes	No		No	Yes	Yes	No	No	No	No	Yes		No
21	One	No	No	No		Yes	No	Yes	No	No	Yes	No	No		No
22	One	No	No	Yes		No	Yes	NA	Yes	No	No	No	Yes		No
23	Two	No	No	No	not famous	No	Yes	NA	No	No	No	No	No	Word too small, boring	No
24	One	No	No	No		Yes	Yes	Yes	Yes	No	No	No	Yes		No
25	One	No	Yes	Yes		No	No	Yes	No	No	Yes	No	No		No
26	One	Yes	No	No		No	Yes	Yes	No	Yes	Yes	No	No		No
27	One	No	No	Yes		No	No	Yes	No	No	Yes	No	No	have number	No
28	One	No	No	Yes		No	No	NA	Yes	No	No	No	No	display more information	No
29	One	No	No	No	only browse	No	No	Yes	No	No	Yes	No	No		No
30	One	No	No	No		Yes	Yes	Yes	No	No	No	Yes	No		No
31	Two	No	Yes	No		No	Yes	No	No	No	Yes	No	No		No
32	One	No	No	No		Yes	No	Yes	No	No	Yes	No	No		No
33	One	Yes	No	No		No	Yes	Yes	No	No	No	No	Yes		No
34	Two	No	No	No	not famous	No	Yes	Yes	No	No	Yes	No	No	Word is smaller	No
35	Two	No	No	No	Other hand nothing to do	No	Yes	Yes	No	No	No	No	No	boring	No
36	One	No	No	No		Yes	No	Yes	No	No	Yes	No	No		No
37	One	No	No	No	No need two hands	No	Yes	Yes	No	No	No	Yes	No		No
38	One	Yes	No	No		No	No	Yes	No	No	No	Yes	No		No
39	One	No	Yes	No		No	Yes	Yes	Yes	No	Yes	No	No		No
40	One	No	No	Yes		No	Yes	Yes	No	No	No	No	No	less colourful	No
41	One	No	No	No	enough	No	Yes	Yes	No	Yes	Yes	No	No		No
42	One	No	No	No		Yes	Yes	Yes	No	No	No	No	No	Can read more information	No
43	One	No	Yes	No		No	Yes	Yes	No	No	Yes	No	No		No
44	One	No	Yes	No		No	Yes	Yes	No	No	Yes	No	No		No
45	One	No	Yes	No		No	No	Yes	No	No	No	No	No	Word layout is no good	No
46	One	No	Yes	No		No	No	Yes	No	No	No	No	No	Boring	No
47	One	Yes	No	No		No	Yes	Yes	No	No	No	No	No	Lack of graphic	No
48	One	No	Yes	No		No	No	Yes	No	No	Yes	No	No		No
49	One	No	Yes	Yes		No	No	Yes	No	No	Yes	No	No		No
50	One	No	No	Yes		No	No	Yes	No	No	No	No	No	too simple	No
51	One	No	Yes	No		No	No	Yes	Yes	No	No	No	No	Too simple, loss location	No
52	One	No	No	No	two hands is difficult	No	Yes	Yes	No	No	No	No	No	colour no good	No
53	Two	No	Yes	No		No	No	Yes	Yes	No	No	Yes	No		No
54	One	No	No	No	softkey	No	Yes	Yes	No	No	No	No	No		Yes
55	Two	Yes	Yes	Yes	pay more attention	No	Yes	Yes	Yes	No	No	No	Yes		No
56	One	No	No	Yes		No	No	Yes	No	No	Yes	No	No		No
57	One	Yes	No	No		No	No	Yes	No	No	Yes	No	No		No
58	One	No	No	Yes		No	Yes	Yes	Yes	No	Yes	No	No	more function, cant expect do what	No
59	One	Yes	Yes	No		No	Yes	Yes	No	No	No	No	No	broing, too much word	No
60	One	No	No	Yes		No	No	Yes	No	No	No	No	Yes		No
61	Two	Yes	No	No		No	No	Yes	No	No	No	No	No	Simple, easy to understand	No

17	18	19	20	21
Mislead instruction (blank) I easy to use after practise I full attention when controling mobile	1	0	3	0
Not clear instruction, Use chinese will be better	0	2	0	0
Mislead by 'Blank' I Problem for 'exit function I soft key is simple	3	0	0	0
Need some time to learn how to use softkey, mislead by scroll bar	3	1	1	1
	0	0	0	0
	1	2	0	4
Should use more colours, should have underline or spacing within group	1	0	0	0
If colour contrat is big, it is eye-catching and easy to control. Now, can't have much prediction for control I The flow is strength forward	3	1	6	0
Want to have chinese display, Mislead by 'Blank', not enought colour	1	4	0	1
Should have Chinese Version	0	2	0	1
	1	0	0	0
Press key too fast, when there is 'blank', make error	0	1	0	0
	1	0	0	0
After enger one item, then exist, should stay in the same position I 'Blank' is misleaded I Should have 'number' key for short cut selection	1	5	0	2
'Blank' is quite mislead, have Chinese is better	1	0	1	0
English version is better because word is smaller, but word is too long	3	2	0	0
'Blank' is mislead, word is too small	1	0	0	2
Layout not clear (too much word)	1	0	1	0
screen layout should be more colourful, 'Blank' is misleaded I try to use number key to 'short cut' to select the item	0	0	1	0
Blank is misleaded	0	1	0	0
Feel quite slow	0	0	0	0
	1	0	0	0
The contract between title and content is not large, the information layout is not clear	1	2	0	2
'Exit' function will let people make mistake easily	0	1	0	0
	0	2	1	1
	0	0	0	0
'Blank' is misleaded	1	1	0	0
Manu order should change	1	0	2	0
One hand control is OK because the choices in softkey is less	1	0	0	0
	1	4	1	2
Feel less messy because no icon	1	0	0	0
Screen layout has too much word, too messy	1	3	0	2
	2	5	1	1
feel quite slow	0	1	1	1
	0	0	1	0
	1	1	0	1
	0	0	0	0
'Blank' cause a great problem and will make a error I Slow	0	0	0	0
Should use more colour to display the word(text)	0	2	0	2
	0	1	1	2
'Blank' is misleading	1	0	0	0
'Blank' is misleading	1	0	0	0
Loading is slow	1	0	0	0
	0	1	1	0
	1	0	0	0
	2	0	1	0
	0	0	2	0
'Blank' is misleaded	0	0	0	0
	0	0	0	0
Should give instruction during waiting	0	0	0	0
	2	0	0	0
	0	4	1	2
	0	0	1	3
When 'back', cannot stay in previous item	2	0	2	1
Should have process bar	0	0	0	0
	1	1	0	0
	0	0	0	0
	0	1	0	0
	1	2	1	1
	0	0	1	1

22
all thing OK
Know which input method at the begining
Usually press wrong key (green or red), dont know how to input number, troublesome, should have 'shift' key for cap letter, need practise
User habit is very important, if someone already like one system, difficult to call other to like Semen phone
Password change to '*', make confusing
Show what input method is using
Dont know how to use input method, change input method also have problem ǀ messy -> press two characters in the same key -> need to wait
Very smooth because he use nokia phone also ǀ Menu have too much selection items
The input is not troublesome
Very easy and clear because it show which input method are using, quicker
Need some time to famous
Word too small, less error because have experience
Should have voice input ǀ very troublesome ǀ should remember login (but not password)
Not clear instruction
default should not be Chinese Input
'*' in password -> can't check
messy for selecting input method ǀ dont know why key will be eaten
string input should only have one type (big cap or small cap only)
Happy with the password system
It is no good to change input method at the begining ǀ Password change to *, cant check
Too many input method in the selection list
Very troublesome since we can check the password
It is good to display input method
Cant expect the password will become *, cant check after input ǀ input is messy, should not have cap letter
not famous with nokia phone
Feel better because it is the second time to do this task, input number-character is good
Should be case insensitive, after input login, there is a long menu which is quite mislead
Password should not ne case sensitive, too many choice in menu bar
Have some problem for changing input method
Input method is not smart enough ǀ * is quite mislead and just think input error ǀ dont know how to input number
Slower than using keyboard
Very troublesome
Very convenient
Cap letter make some problem, system not smart enough ǀ User usually expect there is a 'enter' after inputting login or password ǀ Character input is so troublesome
Step by step -> will not miss any step
Dont expect it is case sensitive ǀ '*' in the password is mislead, should be change to other character which cannot be input
Can know what input method is using
Case sensitive is troublesome ǀ it is goos that login and password are in the same page
Very slow for input ǀ Should not be case sensitive as it is so troublesome
Too many selection item in manu which make confusing
Expect to have 'enter' after login or password ǀ Case sensitive is no good
Use one hand is difficult for input ǀ Input number is difficult
Password change to * is quite mislead ǀ a little bit difficult for input number
Case sensitive for input is quite troublesome ǀ no need to change to * for password because it is difficult to verify, or use number in this case may be better ǀ Dont repear use same key for difficult continuously
Very troublesome, dont know how to input password
Difficult ot change input method
Dont like it at the begining, so troublesome, many step
Dont know how to input Cap Letter
Change to "*" have problem
feel troublesome, need to press few keys for one char, sometimes need to wait
Login have problem, expect have 'enter' key after login
Very messy

23	24	25	26	27	28	29	30	31	32	33
1	1	0	1	1	0	75	45	108	142	16
2	0	0	2	2	0	33	45	183	199	10
0	2	0	0	2	0	84	47	67	63	30
2	2	0	1	2	0	41	84	111	106	23
0	4	0	0	3	0	49	49	56	86	36
1	2	0	0	0	0	51	91	60	49	26
0	3	0	0	1	0	36	33	51	58	36
2	1	0	1	0	1	218	61	62	175	36
1	1	0	0	1	0	80	46	60	97	34
1	4	0	0	3	0	34	52	48	157	38
0	4	0	0	2	0	43	44	70	156	27
3	2	1	1	3	1	46	45	58	64	162
1	6	0	0	1	0	90	52	109	338	24
1	5	0	0	4	0	58	155	71	155	37
1	5	1	0	2	0	87	47	65	266	58
1	4	0	0	2	0	52	52	84	128	28
2	3	1	1	0	0	59	79	54	72	62
1	5	0	0	0	0	46	41	47	120	24
1	1	0	0	2	0	71	29	80	77	25
1	4	1	0	3	0	53	55	101	201	79
1	1	0	0	0	0	27	32	57	73	22
0	4	0	0	1	0	38	62	42	152	25
0	3	0	0	2	0	90	150	166	289	71
0	5	0	0	0	0	29	32	49	168	25
0	4	0	0	3	0	46	56	44	214	30
1	3	0	0	2	0	31	36	64	342	29
1	0	0	0	0	0	48	49	54	105	40
2	1	0	0	2	0	40	57	68	73	28
1	5	1	0	2	0	57	63	49	242	30
1	2	1	0	0	0	75	112	127	90	36
0	0	2	0	1	1	76	47	54	41	92
0	4	0	0	1	0	42	24	44	148	15
99	99	99	99	99	99	60	181	59	69	41
0	3	1	0	1	0	46	97	37	165	38
0	2	1	0	2	0	64	46	58	142	110
2	4	0	1	1	0	62	50	37	134	24
1	5	0	0	3	0	30	52	54	313	27
0	5	0	0	0	0	41	55	61	86	44
0	2	0	0	1	0	36	136	49	102	33
2	1	0	0	3	0	60	35	42	248	30
0	3	0	0	2	0	36	28	38	165	15
0	3	0	1	2	0	43	37	47	187	23
0	4	0	0	2	0	65	40	60	188	23
1	3	2	0	2	0	81	57	74	244	145
0	3	0	0	0	0	38	24	55	62	27
0	5	0	0	2	0	153	103	100	369	34
0	2	0	0	1	0	141	53	61	163	36
1	0	0	0	0	0	28	71	49	93	162
1	6	0	0	0	0	39	78	112	143	22
1	2	0	0	0	0	56	57	68	156	28
1	2	1	0	3	2	46	62	100	98	95
1	3	1	0	1	1	48	144	44	179	78
0	3	0	0	1	0	57	120	37	116	29
0	1	0	0	0	0	169	81	60	63	51
0	2	0	0	0	0	66	32	43	151	23
0	2	0	0	2	0	54	47	68	109	47
0	4	0	0	2	0	39	56	52	206	28
0	4	0	0	2	0	44	64	42	37	43
0	5	0	0	2	0	99	87	53	119	24
0	2	0	0	1	0	53	75	36	129	111

34	35
Not clear instruction (Book ticket), dont't know how to do ǀ 2nd experience is more easy since more time to practise	NA
Response time for pressing key is very slow ǀ mislead by 'blank' ǀ 2nd application will not become famous because they are two different program in different bland phone	NA
Response time is very slow ǀ input data is step-by-step ǀ Scroll bar can tell whether have extra information or not	NA
Very slow, cant bear the response time	NA
	Yes
Not bad but very slow, should have some instruction during loading	NA
Dont know when use 'green key', not clear instruction	NA
If no hint, dont know when to press green key to reserve ǀ Slow, Cant bear the waiting time ǀ when pressback, go to the begining -> problem ǀ He try to use keypad for short cut for selection list	NA
Response time is too long, cant bear	Yes
Dont know already press key or not (no acknowledge), too long response time, can bear the time for moving sit	Yes
Slow, should have some message for loading (Clear instruction)	Yes
Very slow, not enough instruction during processing (dont't know it is loading) ǀ Cant accept the speed for zone, can accpet the speed for moving	No
Layout is clear because it is English ǀ cant guess how to control ǀ too slow	No
Not clear instruction, suddently need to press green key ǀ very slow	No
Difficult to use, very slow, after press back, easy to get loss	Yes
Too slow, dont have any response after key pressing because it is too slow	Yes
Very slow, time delay too long	No
Very slow for the selection, and no response, dont know alreadu press key or not	Yes
cant predict how long for the selection list ǀ click too fast -> afraid miss ǀ very slow for 'chair selection' -> afraid wrong	Yes
Response time is too slow ǀ Cant predict how to use highlight sit	NA
Very slow and dont have instruction	Yes
Slow, but it is quite well to use, should have instruction when it is loading	Yes
The instruction is not enough to teach how to control, too long response time	NA
Response time is very slow, should have instruction for waiting	Yes
Very slow ǀ 'Back' function is not good since need to re-do all the thing ǀ not clear instruction how to control ǀ if he have been instructed, can bear that long waiting time	No
Loading should have instruction	Yes
Time delay -> need sound effect	Yes
Time delay to server is long, should have instruction for waiting	NA
Repeat to press the same key because no instruction during waiting ǀ waiting is acceptable but should have instruction	No
Should back step by step ǀ the waiting time is too slow so that repeat to press the same key	Yes
Response time is too slow, dont accept ǀ scroll system is not good ǀ try to use short cut -> press number key to select	No
It is too slow and just keep on press too many time on a same key because dont know it is work or not, should have instruction for waiting ǀ cant expect how to control	No
Very slow, dont have response, dont know have problem or not	Yes
very slow	NA
slow	Yes
Too slow ǀ quite smooth, step by step	Yes
slow	No
Very slow, difficult to control	Yes
Dont know how to use it (cant expect how to buy ticket)	Yes
Time delay -> especially sitting plan	NA
It is good that softkey dont have label during waiting since it just like the waiting signal	Yes
Very slow and dont have response, this make him just keep on pressing the key in order to have the response	No
Very slow cause difficult to control	No
Back function should not back too early ǀ time delay is too large ǀ Display order of the aviable time have problem	Yes
Very slow, just like no response	No
Time delay, dont have response	Yes
Dont know how to control if I dont give the hints ǀ very slow	No
Very slow, dont know already pressed the key or not, should have instruction during loading	Yes
Very slow	Yes
Time delay -> dont have any response ǀ Scroll down -> 5th item then 1st item, just think it is infinite loop, think there are many items	No
Very slow which easily cause pressing wrong key	Yes
Control is so confused, very slow, should have instruction during loading	Yes
Very slow, but simple, easy to use ǀ Back to too beginning	No
Cant use in moble situation	No
Very slow, should have process bar	Yes
The instruction to select the sits is quite difficult	NA
Very difficult to highlight something, very slow	Yes
Very difficult to control, very slow ǀ 'Back' should not back too early ǀ should have instruction during loading	Yes
Need to redo all the thing if do something wrong	No
If do something wrong, need to re-do all -> troublesome ǀ Very slow	No

Key of each column

1	Ref. No.
2	How many Hands M
3	Convenient M
4	Habit M
5	Phone Shape M
6	Other M
7	NA M
8	difficult walking M
9	can one hand control M
10	sound effect (hints) M
11	Pretty M
12	Clear M
13	Acceptable M
14	Not Acceptable M
15	Other M
16	NA M
17	Remark 1 M
18	Frustration S2T1 M
19	Frustration S2T2 M
20	Extra Help S2T1 M
21	Extra Help S2T2 M
22	String Input M
23	Frustration S3T1 M
24	Frustration S3T2 M
25	Frustration S3T3 M
26	Extra Help S3T1 M
27	Extra Help S3T2 M
28	Extra Help S3T3 M
29	Time1(S2)M
30	Time2(S2)M
31	Time1(S3)M
32	Time2(S3)M
33	Time3(S3)M
34	Overall Control M
35	Nokia User

Data from Questionnaire (Evaluate the Software)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
2	3	2	2	4	2	4	2	2	3	4	3	5	4	4	4	1	4	4	2	2	4	4
3	4	4	4	4	5	4	4	5	4	4	4	3	4	4	4	2	4	3	3	2	3	3
4	4	2	2	3	3	4	3	2	2	3	3	3	4	4	4	3	3	3	4	3	3	4
5	4	5	4	4	3	3	3	2	3	4	4	1	3	2	3	1	4	3	3	4	3	3
6	3	2	4	3	1	4	2	3	2	4	3	2	3	3	4	1	3	4	3	3	4	3
7	3	4	2	3	3	2	3	3	3	4	4	2	4	4	4	2	4	4	3	3	4	4
8	4	4	4	3	3	5	2	3	4	4	3	1	2	2	2	2	4	4	2	2	5	2
9	3	5	4	4	3	3	4	5	4	3	4	3	2	3	2	1	3	2	2	2	3	2
10	3	3	2	3	4	3	3	2	2	3	3	2	3	4	4	2	4	3	3	2	3	2
11	5	5	4	4	5	5	4	4	5	5	4	3	4	4	4	3	4	4	5	4	5	4
12	2	4	3	2	1	2	1	3	1	3	2	3	4	4	5	4	4	3	5	4	4	4
13	5	4	4	5	3	3	3	4	2	3	4	1	3	4	3	1	3	2	4	3	3	3
14	3	2	2	2	3	3	3	1	2	3	3	3	4	4	4	4	3	3	4	4	4	4
15	4	4	4	4	4	4	3	2	2	3	3	1	2	4	3	1	2	1	1	2	3	2
16	2	3	3	1	3	2	1	2	1	2	2	1	2	1	1	1	1	2	2	2	2	2
17	2	4	1	1	2	3	2	1	2	2	2	5	3	4	2	1	3	2	4	2	2	2
18	3	4	2	4	4	3	4	1	2	3	4	3	2	4	4	1	2	3	2	4	4	3
19	4	4	3	4	4	2	3	4	2	4	4	3	2	2	3	1	3	1	2	3	4	3
20	4	3	3	4	4	3	3	2	2	2	2	3	4	5	4	2	4	3	2	2	3	3
21	3	4	2	3	1	4	1	3	1	3	2	2	2	5	4	2	5	3	4	3	4	4
22	4	5	2	4	4	4	3	4	1	4	4	2	2	4	3	1	4	2	2	2	4	3
23	4	3	2	3	3	1	1	2	1	2	2	3	4	4	4	2	4	2	4	4	4	4
24	2	3	2	3	4	3	3	3	4	3	1	2	1	2	3	2	2	2	2	1	3	2
25	3	2	2	2	2	3	2	2	1	4	2	3	4	4	4	2	4	4	3	2	4	3
26	3	3	3	4	4	4	3	4	3	5	4	5	4	3	3	3	2	3	2	2	3	3
27	5	4	3	5	3	4	3	3	3	5	3	4	4	4	4	2	4	3	4	3	4	4
28	3	4	2	3	3	2	3	3	3	5	3	1	5	4	4	5	3	3	4	3	5	4
29	5	4	4	5	3	3	3	3	3	4	3	2	4	4	3	4	4	4	5	5	5	4
30	3	3	3	2	4	2	2	1	2	4	2	3	3	3	4	1	4	3	3	2	4	3
31	3	3	2	2	3	3	3	2	4	3	3	3	3	2	2	2	3	2	4	4	3	3
32	5	4	3	4	2	2	2	3	2	4	4	3	4	4	3	1	2	2	3	3	4	3
33	4	5	3	3	4	3	2	4	2	3	3	5	3	2	3	2	3	2	3	3	4	3
34	3	3	2	3	2	3	3	2	2	3	3	2	2	4	3	1	2	2	3	3	3	3
35	3	3	2	2	2	2	2	2	3	2	2	3	2	2	3	2	3	2	2	2	3	3
36	3	3	2	2	2	3	1	2	3	3	1	3	3	3	4	2	3	2	3	4	3	3
37	3	4	2	2	4	2	3	2	2	3	2	3	4	4	3	1	3	1	4	4	1	2
38	3	3	2	3	2	2	2	1	3	3	2	3	3	3	4	2	2	2	3	2	3	3
39	5	4	3	3	3	3	4	2	3	4	3	4	3	3	2	1	2	2	2	3	4	2
40	2	3	2	3	2	3	3	1	1	3	3	3	2	4	4	2	3	4	2	4	3	4
41	4	3	2	2	5	4	4	2	2	4	3	5	4	3	4	3	4	3	5	2	4	3
42	4	3	4	4	3	5	4	5	3	4	4	4	4	5	4	3	4	4	3	3	4	4
43	3	3	2	3	3	3	2	2	3	3	3	3	4	4	3	2	3	3	3	3	4	3
44	3	3	4	4	2	4	3	2	3	4	4	2	4	4	3	1	4	2	4	3	4	3
45	4	4	1	3	3	3	3	1	1	3	3	4	4	2	2	1	3	1	2	4	4	3
46	2	3	4	3	3	4	4	3	3	4	3	3	3	4	2	1	4	4	2	3	3	3
47	4	3	3	4	4	3	3	2	3	4	4	2	1	1	2	1	2	2	1	2	3	2
48	3	3	4	4	4	4	4	2	4	3	3	1	3	5	3	2	4	2	3	4	3	4
49	3	4	2	3	4	1	2	2	1	4	2	3	5	4	4	1	2	2	2	1	5	2
50	4	4	3	3	3	3	3	2	3	3	3	2	3	2	2	1	2	2	3	3	3	2
51	4	5	3	5	4	2	4	4	4	5	4	1	2	3	4	3	2	3	1	2	5	3
52	3	4	4	3	2	2	3	2	2	2	3	1	4	4	2	3	2	4	3	2	3	3
53	3	4	2	3	4	4	3	4	4	2	4	2	2	2	3	3	4	3	2	4	5	3
54	2	4	2	2	2	3	3	2	2	4	2	1	1	2	3	1	3	4	3	3	3	3
55	4	4	4	3	3	3	3	3	3	3	3	3	2	2	3	2	2	2	2	2	2	2
56	3	4	2	3	3	3	3	2	4	4	4	4	4	4	4	3	4	3	4	4	4	4
57	4	5	3	3	4	4	4	3	3	3	3	3	4	4	3	3	4	3	3	4	4	4
58	3	4	3	3	2	3	2	3	4	3	3	2	2	2	2	2	2	2	3	2	3	3
59	3	4	2	3	2	4	3	2	3	3	3	1	2	3	3	2	2	2	3	3	3	3
60	2	4	2	2	1	2	2	3	3	3	3	1	1	2	3	1	2	2	2	2	3	3
61	3	4	4	4	4	4	3	3	3	3	4	3	3	3	4	2	3	2	2	3	3	3

Key of each column

1	Ref. No.
2	sound effect T
3	screen layout T
4	input control T
5	feeling T
6	waiting time T
7	inputting time T
8	speed of use T
9	clear instruction T
10	accuracy T
11	task completeness T
12	overall rating T
13	sound effect M
14	screen layout M
15	input control M
16	feeling M
17	waiting time M
18	inputting time M
19	speed of use M
20	clear instruction M
21	accuracy M
22	task completeness M
23	overall rating M

Data from Question and Answer Section

1	2	3	4	5	6	7	8	9	10	11
2	Same	NA	NA	NA	NA	Same	4	S3T3	5	
3	Same	Same	Travel	Travel	Travel	Travel	6		5	
4	Movie	NA	Same	Movie	Movie	Movie	4	S3T3	5	
5	NA	Travel	Travel	Travel	Travel	Travel	4	S3T3	4	S3T3
6	Same	Movie	Travel	Movie	Travel	Travel	5		5	
7	NA	Same	Same	Travel	NA	NA	4	S2T2	5	
8	Travel	NA	Same	Travel	Movie	Travel	5		5	
9	Travel	NA	Travel	Travel	Travel	Travel	5		5	
10	NA	Movie	Travel	Travel	Movie	Travel	5		5	
11	Same	NA	Travel	NA	Same	Travel	5		5	
12	Movie	NA	Same	Movie	Movie	Movie	5		5	
13	Movie	NA	Travel	Travel	Movie	Travel	5		5	
14	Same	NA	Same	Movie	Movie	Movie	5		5	
15	Movie	NA	Same	Travel	Travel	Travel	4	S2T2	5	
16	Same	NA	Same	Travel	Travel	Travel	5		5	
17	Movie	NA	Movie	Movie	Travel	Movie	5		5	
18	Movie	NA	Same	Travel	Same	Movie	5		5	
19	Same	NA	Same	Travel	Travel	Travel	5		5	
20	Movie	NA	Same	Travel	NA	Movie	4	S3T1	5	
21	Same	NA	Same	Movie	Travel	Movie	5		5	
22	Travel	NA	Travel	Travel	Movie	Movie	5		5	
23	Same	NA	Same	Travel	Movie	Movie	5		5	
24	Travel	NA	Travel	Travel	Travel	Travel	5		4	S2T2
25	Movie	NA	Same	Movie	Movie	Movie	4	S3T3	5	
26	Same	Movie	Same	Movie	Travel	Same	5		4	S2T2
27	Movie	NA	Travel	Movie	Movie	Movie	5		5	
28	NA	Movie	Travel	Movie	Same	Movie	5		5	
29	NA	Movie	Same	Movie	Travel	Movie	5		5	
30	Movie	NA	Same	Movie	Movie	Movie	5		5	
31	Same	Movie	Same	Movie	Travel	Movie	4	S2T2	4	S3T2
32	Same	Same	Travel	Movie	Travel	Movie	5		5	
33	Same	Same	Same	Travel	Movie	Movie	5		5	
34	Movie	Movie	Same	Travel	Movie	Same	5		5	
35	NA	Movie	Travel	Travel	Movie	Movie	5		5	
36	NA	Same	Same	Movie	Movie	Movie	3	S2T2,S3T3	5	
37	Same	Movie	Movie	Travel	Travel	Same	4	S2T2	5	
38	Same	Movie	Same	Travel	Movie	Travel	5		5	
39	Travel	NA	Travel	Travel	Same	Same	5		5	
40	Same	Movie	Same	Travel	Movie	Movie	4	S2T2	5	
41	Travel	Movie	Same	Movie	Travel	Movie	4		5	
42	Same	Same	Same	Movie	Movie	Same	5		5	
43	Same	Movie	Same	Movie	Same	Movie	5		5	
44	NA	Same	Same	Same	Movie	Same	4	S2T2	5	
45	Same	Movie	Travel	Travel	Movie	Movie	4	S2T2	5	
46	Same	Movie	Same	Travel	Movie	Same	5		5	
47	Same	Travel	Travel	Travel	Travel	Travel	5		5	
48	Movie	Movie	Same	Movie	Movie	Travel	5		5	
49	Movie	NA	Same	Movie	Movie	Movie	5		5	
50	Same	Same	Travel	Travel	Travel	Movie	4	S2T3	5	
51	Travel	Same	Travel	NA	Movie	Travel	5		5	
52	Same	Movie	Same	Travel	Movie	Travel	5		5	
53	Movie	Movie	Same	Travel	Same	Travel	5		5	
54	Same	Movie	Same	Travel	Travel	Movie	5		5	
55	Travel	Travel	Travel	Travel	Travel	Travel	5		5	
56	Same	Movie	Same	Movie	Travel	Movie	5		5	
57	Same	NA	Travel	Travel	Travel	Movie	5		5	
58	Movie	Movie	Travel	Movie	Movie	Travel	5		5	
59	Movie	Movie	Travel	Travel	Same	Movie	4	S2T2	5	
60	Movie	Same	Same	Travel	Travel	Travel	5		5	
61	NA	Same	Same	Travel	Movie	Travel	4	S2T2	5	

Key of each column

1	Ref. No.
2	walking
3	one hand control
4	sound effect
5	screen layout
6	input
7	overall control
8	Task Completeness T
9	Incompleted List T
10	Task Completeness M
11	Incompleted List M

Data from Questionnaire (Evaluate Users Background)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2	Sim	Male	19	Yes	4	No	No	Yes	Not more than five time	No	No	No	No	Yes	No	No	Yes	Yes	No	No
3	Nokia	Male	18	Yes	4	No	Yes	Yes	Once per week	Yes	Yes	No	Yes	Yes	No	No	Yes	Yes	No	No
4	Sim	Male	20	Yes	4	No	Yes	Yes	Not more than five time	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
5	Nokia	Male	20	Yes	3.5	Yes	Yes	Yes	Not more than five time	Yes	Yes	No	No	No	No	No	No	Yes	No	No
6	Sim	Male	20	Yes	4	No	No	Yes	One per day	Yes	Yes	No	No	No	No	No	No	Yes	No	No
7	Nokia	Male	19	Yes	5	Yes	Yes	Yes	Once per week	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	No	No
8	Sim	Male	19	Yes	3.5	Yes	Yes	Yes	Once per month	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	No
9	Nokia	Male	22	Yes	3.25	No	No	Yes	Once per month	Yes	Yes	No	No	No	No	No	No	No	No	No
10	Sim	Male	20	Yes	4.5	No	Yes	Yes	Not more than five time	No	No	No	No	No	No	Yes	Yes	No	No	Yes
11	Nokia	Female	20	Yes	3	No	No	Yes	One per day	Yes	No	No	Yes	Yes	No	Yes	No	Yes	No	No
12	Sim	Male	20	Yes	7	No	No	Yes	Once per week	No	No	No	Yes	Yes	No	No	Yes	Yes	Yes	No
13	Nokia	Male	20	Yes	2	No	No	No	.	No	No	No	No	No	No	Yes	No	Yes	No	No
14	Sim	Female	24	Yes	5	No	Yes	Yes	Not more than five time	Yes	Yes	No	No	No	No	No	No	No	No	No
15	Nokia	Male	24	Yes	6	No	No	Yes	Once per week	No	Yes	No	Yes	No	No	No	Yes	Yes	Yes	Yes
16	Sim	Male	23	Yes	6	No	No	Yes	Not more than five time	No	Yes	No	No	No	No	No	No	No	No	No
17	Sim	Male	22	Yes	5	No	No	No	.	Yes	Yes	No	No	No	No	No	No	No	No	No
18	Nokia	Male	19	Yes	4	No	No	Yes	Once per week	Yes	Yes	Yes	No	Yes	No	No	No	Yes	Yes	Yes
19	Sim	Female	21	Yes	3.3	No	No	Yes	More than once per day	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	No	No
20	Nokia	Male	23	Yes	4	No	No	Yes	Once per week	Yes	Yes	No	No	No	No	No	No	No	No	No
21	Sim	Male	22	Yes	5	No	No	Yes	Once per month	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes	No
22	Nokia	Male	23	Yes	5	No	Yes	Yes	Once per week	Yes	Yes	No	No	No	No	No	No	No	No	No
23	Sim	Female	23	Yes	4	No	No	Yes	Once per week	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	No	No
24	Nokia	Male	37	Yes	7	Yes	No	No	.	Yes	Yes	No	No	No	No	No	No	No	No	No
25	Sim	Female	23	Yes	5	No	No	Yes	More than once per day	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
26	Nokia	Female	21	Yes	5	No	Yes	Yes	Once per month	No	No	No	No	No	No	No	No	No	No	No
27	Sim	Female	21	Yes	5	No	No	Yes	Once per month	No	No	No	No	No	No	No	No	No	No	No
28	Sim	Male	21	Yes	2.5	No	No	Yes	Not more than five time	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	No	No
29	Nokia	Female	23	Yes	4	No	No	Yes	Once per month	Yes	Yes	No	No	No	No	No	No	No	No	No
30	Sim	Female	23	Yes	4	No	Yes	Yes	Not more than five time	Yes	Yes	No	No	No	No	No	No	No	No	No
31	Nokia	Female	21	Yes	5	No	No	Yes	More than once per day	Yes	No	No	No	No	No	No	No	No	No	No
32	Sim	Male	19	Yes	4	No	No	Yes	Once per month	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	No	Yes
33	Nokia	Female	19	Yes	0.25	No	No	Yes	More than once per day	No	Yes	No	No	No	No	No	No	No	No	No
34	Sim	Female	22	Yes	4.5	No	No	Yes	One per day	Yes	No	No	No	No	No	No	No	No	No	Yes
35	Nokia	Female	19	Yes	1.75	No	No	Yes	Once per month	No	No	No	No	No	No	No	No	No	No	No
36	Sim	Female	18	Yes	3	No	No	Yes	Once per week	Yes	Yes	No	No	No	No	No	No	Yes	No	No
37	Nokia	Female	24	Yes	7	No	Yes	Yes	More than once per day	No	Yes	No	Yes	No	No	Yes	Yes	No	No	No
38	Sim	Female	19	Yes	4.1	No	No	Yes	Not more than five time	No	No	No	Yes	No	No	No	Yes	Yes	Yes	Yes
39	Nokia	Female	20	Yes	5	No	Yes	Yes	One per day	No	Yes	No	Yes	Yes	No	Yes	Yes	No	No	No
40	Sim	Male	18	Yes	2.5	No	No	Yes	More than once per day	No	No	No	No	No	No	No	No	No	No	No
41	Nokia	Male	21	Yes	3	No	No	Yes	Once per month	No	No	No	Yes	No	No	Yes	Yes	Yes	No	No
42	Sim	Female	24	Yes	5	Yes	No	Yes	More than once per day	Yes	Yes	No	No	No	No	No	No	Yes	No	No
43	Nokia	Female	20	Yes	5	No	No	Yes	Once per week	No	Yes	No	No	No	No	No	Yes	Yes	No	No
44	Sim	Male	19	Yes	5	Yes	No	Yes	Once per week	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	No
45	Nokia	Male	24	Yes	5	No	No	Yes	Once per month	Yes	Yes	No	No	No	No	No	No	No	No	No
46	Sim	Male	18	Yes	3.5	Yes	Yes	Yes	Once per month	Yes	Yes	No	No	No	No	No	No	No	No	No
47	Nokia	Male	23	Yes	5	No	No	No	.	Yes	Yes	No	No	No	No	No	No	Yes	No	No
48	Sim	Female	21	Yes	4	Yes	No	Yes	More than once per day	No	Yes	No	No	No	No	No	No	No	No	No
49	Nokia	Male	25	Yes	6	Yes	No	Yes	Not more than five time	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	Yes	No
50	Nokia	Male	21	Yes	4	No	No	Yes	More than once per day	No	Yes	No	No	No	No	No	No	No	No	No
51	Nokia	Male	21	Yes	5.1	Yes	No	Yes	Once per month	No	Yes	No	Yes	No	No	No	Yes	Yes	Yes	Yes
52	Sim	Female	20	Yes	3.5	No	No	No	.	Yes	Yes	No	No	No	No	Yes	Yes	Yes	No	No
53	Nokia	Female	20	Yes	4	No	No	Yes	Not more than five time	Yes	No	No	Yes	No	No	No	No	No	No	No
54	Sim	Female	24	Yes	4	No	No	Yes	More than once per day	No	Yes	No	Yes	No	No	Yes	Yes	Yes	No	No
55	Nokia	Female	28	Yes	3.25	No	No	Yes	Once per week	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No
56	Sim	Female	25	Yes	4	Yes	No	Yes	Not more than five time	Yes	Yes	Yes	Yes	No	No	No	No	No	No	Yes
57	Nokia	Female	22	Yes	5	No	Yes	Yes	Once per week	No	Yes	No	Yes	Yes	No	No	Yes	No	Yes	No
58	Sim	Female	24	Yes	5	No	No	Yes	Once per month	No	No	No	No	No	No	No	No	No	No	No
59	Nokia	Female	22	Yes	5	No	No	Yes	Once per month	No	Yes	No	Yes	No	No	No	Yes	Yes	No	No
60	Sim	Female	20	Yes	5	No	No	Yes	Once per week	No	No	No	No	No	No	No	No	No	No	No
61	Nokia	Female	21	Yes	3	Yes	No	Yes	Once per month	No	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No

22	23	24	25	26	27	28	29	30	31	32
Yes	No	No	No	Yes	No	4	5	3	2	4
Yes	No	Yes	Yes	Yes	Yes	4	7	4	3	4
Yes	Yes	Yes	Yes	Yes	Yes	3	.	4	4	4
Yes	No	No	Yes	Yes	Yes	4	3	4	4	3
Yes	No	Yes	Yes	No	Yes	4	5	4	4	4
Yes	No	No	Yes	No	Yes	2	5	4	4	4
Yes	No	No	Yes	No	Yes	4	4	4	4	5
Yes	No	Yes	Yes	Yes	Yes	16	11	3	4	3
Yes	Yes	No	Yes	No	No	5	6	4	5	4
Yes	No	No	Yes	No	Yes	3	4	5	5	4
No	No	No	Yes	Yes	Yes	5	10	4	4	4
No	No	Yes	Yes	No	No	.	10	5	4	4
Yes	No	No	Yes	No	Yes	3	3	5	4	4
Yes	No	No	Yes	No	Yes	5	8	5	5	5
No	No	No	Yes	Yes	No	2	5	5	5	2
No	No	No	Yes	No	No	7	4	5	5	4
Yes	No	No	Yes	Yes	Yes	3	10	4	5	5
Yes	No	Yes	Yes	Yes	Yes	3	6	5	4	5
Yes	No	No	Yes	No	Yes	1	10	2	3	3
Yes	No	No	Yes	No	Yes	4	6	4	5	3
Yes	No	No	Yes	No	Yes	3	3	5	5	4
No	No	Yes	Yes	Yes	Yes	20	8	5	5	5
Yes	No	No	Yes	Yes	Yes	7	15	5	5	5
Yes	Yes	Yes	Yes	Yes	Yes	2	.	5	5	3
Yes	No	No	Yes	No	Yes	4	4	4	4	3
Yes	No	No	Yes	Yes	Yes	3	5	5	4	4
Yes	No	Yes	Yes	Yes	Yes	5	5	4	3	4
Yes	No	No	Yes	Yes	Yes	3	5	5	4	5
No	No	No	Yes	No	Yes	3	6	5	5	5
Yes	No	No	Yes	No	Yes	4	8	3	2	3
Yes	Yes	No	Yes	Yes	Yes	.	9	4	4	3
No	No	No	Yes	No	Yes	3	3	3	4	4
Yes	No	No	Yes	No	Yes	7	2	4	3	3
Yes	No	Yes	Yes	No	Yes	4	7	3	3	3
Yes	No	Yes	Yes	Yes	Yes	3	10	5	5	4
Yes	Yes	No	No	No	No	.	5	4	4	4
Yes	No	No	Yes	No	Yes	3	2	3	3	4
Yes	No	Yes	Yes	Yes	Yes	3	8	5	5	4
Yes	No	No	Yes	Yes	Yes	16	15	3	3	4
Yes	No	Yes	Yes	Yes	Yes	5	5	5	4	4
No	No	No	Yes	No	Yes	3	8	4	3	4
Yes	No	No	Yes	No	Yes	10	5	4	3	4
Yes	No	Yes	Yes	Yes	Yes	2	10	4	5	5
No	No	No	No	No	Yes	16	6	4	5	3
Yes	No	Yes	Yes	No	Yes	2	15	4	3	3
No	No	No	Yes	No	Yes	5	3	4	5	4
No	No	No	Yes	No	Yes	3	3	4	4	4
Yes	Yes	No	Yes	No	Yes	4	10	4	3	3
No	No	No	Yes	Yes	Yes	4	15	5	5	5
Yes	No	Yes	Yes	Yes	Yes	.	10	4	4	5
No	No	No	No	No	No	3	1	3	5	3
Yes	No	No	Yes	No	Yes	5	5	5	5	5
No	No	Yes	Yes	Yes	No	4	2	5	5	5
Yes	No	No	Yes	No	Yes	4	8	5	4	4
Yes	No	No	Yes	Yes	Yes	6	2	4	4	3
Yes	No	No	Yes	No	Yes	.	7	5	4	4
No	No	No	Yes	No	No	5	3	4	5	4
Yes	Yes	Yes	Yes	Yes	Yes	7	11	4	3	4
Yes	No	Yes	Yes	No	Yes	4	.	5	4	4
Yes	No	Yes	Yes	Yes	Yes	3	4	4	4	3

Key of each column

1	Ref. No.
2	First application used
3	Sex
4	Age
5	mobile phone user
6	how long for m-user
7	Use WAP
8	Use JAVA software
9	send SMS
10	How often sending SMS
11	IT background
12	heard 3G?
13	Bank Transaction
14	Buy cinema ticket
15	Pay bills
16	Buy Things
17	Check travel info
18	Check weather
19	Read news
20	Check Movie
21	ICQ
22	Mini games
23	Online games
24	Organizer
25	Phone book
26	Reminder
27	Calculator
28	No. of colour
29	No. of software
30	Effectiveness
31	Efficiency
32	Satisfaction